LINEAR OPTIMAL CONTROL

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

BRIAN D. O. ANDERSON
Professor of Electrical Engineering
University of Newcastle
New South Wales, Australia

JOHN B. MOORE
Associate Professor
Department of Electrical Engineering
University of Newcastle

LINEAR OPTIMAL CONTROL

PRENTICE-HALL, INC.
Englewood Cliffs, New Jersey

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

13–536870–7

Library of Congress Catalog Card No. 71–137984

Printed in the United States of America

PREFACE

Despite the development of a now vast body of knowledge known as modern control theory, and despite some spectacular applications of this theory to practical situations, it is quite clear that much of the theory has yet to find application, and many practical control problems have yet to find a theory which will successfully deal with them. No book of course can remedy the situation at this time. But the aim of this book is to construct one of many bridges that are still required for the student and practicing control engineer between the familiar classical control results and those of modern control theory. It attempts to do so by consistently adopting the viewpoints that

1. many modern control results have interpretation in terms of classical control results;
2. many modern control results do have practical engineering significance, as distinct from applied mathematical significance.

As a consequence, linear systems are very heavily emphasized, and, indeed, the discussion of nonlinear systems is essentially restricted to two classes: systems which should be linear, but unhappily are not; and systems which are linear but for the intentional use of relays. Also as a consequence of this approach, discussion of some results deemed fundamental in the general theory of optimal control has been kept to the barest minimum, thereby allowing emphasis on those particular optimal control results having application to linear systems. It may therefore seem strange to present a book on optimal control which does not discuss the Pontryagin Maximum Principle, but it is nonetheless consistent with the general aims of the book.

Although the selection of the material for the book has not been governed

by the idea of locating the optimal control theory of linear systems within the broader framework of optimal control theory per se, it has been governed by the aim of presenting results of linear optimal control theory interesting from an engineering point of view, consistent with the ability of students to follow the material. This has not meant restricting the choice of material presented to that covered in other books; indeed a good many of the ideas discussed have appeared only in technical papers.

For the most part, continuous time systems are treated, and a good deal more of the discussion is on time-invariant than is on time-varying systems. Infinite-time optimization problems for time-varying systems involve concepts such as uniform complete controllability, which the authors consider to be in the nature of advanced rather than core material, and accordingly discussion of such material is kept to a minimum. For completeness, some mention is also made of discrete-time systems, but it seemed to us that any extended discussion of discrete-time systems would involve undue repetition.

The text is aimed at the first or later year graduate student. The background assumed of any reader is, first, an elementary control course, covering such notions as transfer functions, Nyquist plots, root locus, etc., second, an elementary introduction to the state-space description of linear systems and the dual notions of complete controllability and complete observability, and third, an elementary introduction to linear algebra. However, exposure to a prior or concurrent course in optimal control is not assumed. For students who have had a prior course, or are taking concurrently a course in the general theory of optimal control or a specific aspect of the discipline such as time-optimal systems, a course based on this book will still provide in-depth knowledge of an important area of optimal control.

Besides an introductory chapter and a final chapter on computational aspects of optimal linear system design, the book contains three major parts. The first of these outlines the basic theory of the linear regulator, for time-invariant and time-varying systems, emphasizing the former. The actual derivation of the optimal control law is via the Hamilton-Jacobi equation which is introduced using the Principle of Optimality. The infinite-time problem is considered, with the introduction of exponential weighting in the performance index used for time-invariant design as a novel feature. The second major part of the book outlines the engineering properties of the regulator, and attempts to give the reader a feeling for the use of the optimal linear regulator theory as a design tool. Degree of stability, phase and gain margin, tolerance of time delay, effect of nonlinearities, introduction of relays, design to achieve prescribed closed-loop poles, various sensitivity problems, state estimation and design of practical controllers are all considered. The third major part of the book discusses extensions to the servomechanism problem, to the situation where the derivative of the control

may be limited (leading to dynamic feedback controllers) or the control itself may be limited in amplitude (leading to feedback controllers containing relays), and to recent results on output, as distinct from state feedback. Material on discrete time systems and additional material on time-varying continuous systems is also presented. The final part of the book, consisting of one chapter only, discusses approaches to the solution of Riccati equations, including approximate solution procedures based on singular perturbation theory. Appendices summarizing matrix theory and linear system theory results relevant to the material of the book are also included.

Readers who have been introduced to the regulator problem elsewhere may find section 3 of chapter 3 a convenient starting point, unless review of the earlier material is required.

We would like to emphasize that the manuscript was compiled as a truly joint effort; it would be difficult to distinguish completely who wrote what section and whose ideas were involved at each point in the development of the material. Both of us were surprised at the fact that working together we could achieve far more than either of us working independently, and we are thankful for the personal enrichment to our lives from the experience of working together.

In listing acknowledgments, our former teachers Robert Newcomb and Dragoslav Ŝiljak come immediately to mind as do our graduate students Peter Moylan and Konrad Hitz. In knowing these people each of us has learned that a teacher-student relationship can be infinitely more worthwhile than the usual connotation of the words implies. We appreciate the direct help given by our students as we also appreciate the work done by Brian Thomas in drafting the various diagrams, and Sue Dorahy, Pam O'Sullivan and Lorraine Youngberry for typing the manuscript. We are happy to acknowledge the financial support of our research by the Australian Research Grants Committee. We mention our families—Dianne and Elizabeth Anderson, and Jan and Kevin Moore because they are a part of us. We mention all these various names along with our own in recognition of the world of people out of whom the world of ideas is born and for whom it exists.

Brian D. O. Anderson
John B. Moore

CONTENTS

LINEAR OPTIMAL CONTROL

PART I

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 LINEAR OPTIMAL CONTROL

The methods and techniques of what is now known as "classical control" will be familiar to most readers. In the main, the systems or plants that can be considered by using classical control ideas are linear, time invariant, and have a single input and a single output. The primary aim of the designer using classical control design methods is to stabilize a plant, whereas secondary aims may involve obtaining a certain transient response, bandwidth, steady state error, and so on. The designer's methods are a combination of analytical ones (e.g., Laplace transform, Routh test), graphical ones (e.g., Nyquist plots, Nichols charts), and a good deal of empirically based knowledge (e.g., a certain class of compensator works satisfactorily for a certain class of plant). For high-order systems, multiple-input systems, or systems that do not possess the properties usually assumed in the classical control approach, the designer's ingenuity is generally the limiting factor in achieving a satisfactory design.

Two of the main aims of modern, as opposed to classical, control are to deempiricize control system design and to present solutions to a much wider class of control problems than classical control can tackle. One of the major ways modern control sets out to achieve these aims is by providing an array of analytical design procedures that lessen the load of the design task

on the designer's ingenuity and locate more of the load upon his mathematical ability and on the computational machines used in actually carrying out the design.

Optimal control is one particular branch of modern control that sets out to provide analytical designs of a specially appealing type. The system which is the end result of an optimal design is not supposed merely to be stable, have a certain bandwidth, or satisfy any one of the desirable constraints associated with classical control, but it is supposed to be the *best* possible system of a particular type—hence, the word optimal. If it is both optimal and possesses a number of the properties that classical control suggests are desirable, so much the better.

Linear optimal control is a special sort of optimal control. The plant that is controlled is assumed linear, and the controller, the device which generates the optimal control, is constrained to be linear. That is, its output, the optimal control, is supposed to depend linearly on its input, which will consist of quantities derived from measurements on the plant. Of course, one may well ask: Why linear optimal control, as opposed simply to optimal control? A number of justifications may be advanced—for example, many engineering plants are linear prior to addition of a controller to them; a linear controller is simple to implement physically, and will frequently suffice.

Other advantages of optimal control, when it is specifically linear, follow.

1. Many optimal control problems do not have computable solutions, or they have solutions that may only be obtained with a great deal of computing effort. By contrast, nearly all linear optimal control problems have readily computable solutions.
2. Linear optimal control results may be applied to nonlinear systems operating on a small signal basis. More precisely, suppose an optimal control has been developed for some nonlinear system with the assumption that this system will start in a certain initial state. Suppose, however, that the system starts in a slightly different initial state, for which there exists some other optimal control. Then a first approximation to the difference between the two optimal controls may normally be derived, if desired, by solving a linear optimal control problem (with all its attendant computational advantages). This holds independently of the criterion for optimality for the nonlinear system. (Since this topic will not be discussed anywhere in this book, we list the two references [1] and [2] that outline this important result.†)
3. The computational procedures required for linear optimal design may often be carried over to nonlinear optimal problems. For example,

†References are located at the end of each chapter.

the nonlinear optimal design procedures based on the theory of the second variation [1–3] and quasilinearization [3, 4] consist of computational algorithms replacing the nonlinear problem by a sequence of linear problems.

4. Linear optimal control designs turn out to possess a number of properties, other than simply optimality, which classical control suggests are attractive. Examples of such properties are good gain margin and phase margin, and good tolerance of nonlinearities. This latter property suggests that controller design for nonlinear systems may sometimes be achieved by designing with the assumption that the system is linear (even though this may not be a good approximation), and by relying on the fact that an optimally designed linear system can tolerate nonlinearities—actually quite large ones—without impairment of all its desirable properties. Hence, linear optimal design methods are in some ways applicable to nonlinear systems.
5. Linear optimal control provides a framework for the unified study of the control problems studied via classical methods. At the same time, it vastly extends the class of systems for which control designs may be achieved.

1.2 ABOUT THIS BOOK IN PARTICULAR

This is not a book on optimal control, but a book on linear optimal control. Accordingly, it reflects very little of the techniques or results of general optimal control. Rather, we study a basic problem of linear optimal control, the "regulator problem," and attempt to relate mathematically all other problems discussed to this one problem. If the reader masters the mathematics of the regulator problem, he should find most of the remainder of the mathematics relatively easy going. (Those familiar with the standard regulator and its derivation may bypass Chapter 2, Sec. 2.1 through Chapter 3, Sec. 3.3. Those who wish to avoid the mathematics leading to regulator results in a first reading may bypass Chapter 2, Sec. 2.2 through Chapter 3, Sec. 3.3.)

The fact that we attempt to set up mathematical relations between the regulator problem and the other problems considered does not mean that we seek, or should seek, physical or engineering relations between the regulator problem and other problems. Indeed, these will not be there, and even the initial mathematical statements of some problems will often not suggest their association with the regulator problem.

We aim to analyze the engineering properties of the solution to the problems presented. We thus note the various connections to classical con-

trol results and ideas, which, in view of their empirical origins, are often best for assessing a practical design, as distinct from arriving at this design.

1.3 PART AND CHAPTER OUTLINE

In this section, we briefly discuss the breakdown of the book into parts and chapters. There are five parts, listed below with brief comments.

Part I—Introduction. This part is simply the introductory first chapter.

Part II—Basic theory of the optimal regulator. These chapters serve to introduce the linear regulator problem and to set up the basic mathematical results associated with it. Chapter 2 sets up the problem, by translating into mathematical terms the physical requirements on a regulator. It introduces the Hamilton–Jacobi equation as a device for solving optimal control problems, and then uses this equation to obtain a solution for problems where performance over a finite (as opposed to infinite) time interval is of interest. The infinite-time interval problem is considered in Chapter 3, which includes stability properties of the optimal regulators. Chapter 4 shows how to achieve a regulator design with a prescribed degree of stability.

Part III—Properties and application of the optimal regulator. The aim of this part is twofold. First, it derives and discusses a number of engineering properties of the linear optimal regulator, and, second, it discusses the engineering implementation of the regulator. The main purpose of Chapter 5 is to derive some basic frequency domain formulas and to use these to deduce from Nyquist plots properties of optimal systems involving gain margin, etc. In this chapter, the problem is also considered of designing optimal systems with prescribed closed-loop poles. In Chapter 6, an examination is made of the effect of introducing nonlinearities, including relays, into optimal systems. The main point examined is the effect of the nonlinearities on the system stability. Chapter 7 is mainly concerned with the effect of plant parameter variations in optimal systems, and studies the effect using modern control ideas as well as the classical notion of the return difference. There is also further discussion in Chapter 7 of the design of optimal systems with prescribed closed-loop poles. Chapter 8 is devoted to the problem of state estimation; implementation of optimal control laws generally requires the feeding back of some function of the plant state vector, which may need to be estimated from the plant input and output if it is not directly measurable. The discussions of Chapter 8 include estimators that operate optimally in the presence of noise, the design of such estimators being achieved via solution of an optimal regulator problem. The purpose of Chapter 9 is to tie

the estimation procedures of Chapter 8 with the optimal control results of earlier chapters so as to achieve controllers of some engineering utility. Attention is paid to simplification of the structure of these controllers.

Part IV—Extensions to more complex problems. In this part, the aim is to use the regulator results to solve a number of other linear optimal control problems of engineering interest. Chapter 10 considers problems resulting in controllers using proportional-plus-integral state feedback. Chapter 11 considers various versions of the classical servomechanism problem. Chapter 12 considers problems when there is an upper bound on the magnitude of the control; the controllers here become dual mode, with one mode—the linear one—computable using the regulator theory. Next, Chapter 13 considers problems where only the plant output is available for use in a nondynamic controller, as well as other optimal problems that include controller constraints. Such problems are often referred to as suboptimal problems. Chapter 14 contains a very brief discussion of discrete time systems, and continuous time-varying systems on an infinite-time interval.

Part V—Computational aspects. This part—Chapter 15—discusses some of the computational difficulties involved in carrying out an optimal control design. Various techniques are given for finding transient and steady state solutions to an equation, the matrix Riccati equation, occurring constantly in linear design. Approximate solutions are discussed, as well as a description of situations in which these approximate solutions are applicable.

Appendices. Results in matrix theory and linear system theory relevant to the material in the book are summarized in the appendices.

REFERENCES

[1] Breakwell, J. V., J. L. Speyer, and A. E. Bryson, "Optimization and Control of Nonlinear Systems Using the Second Variation," *SIAM J. Control*, Vol. 1, No. 2, 1963, pp. 193–223.

[2] Kelley, H. J., "Guidance Theory and Extremal Fields," *IRE Trans. Auto. Control*, Vol. AC-7, No. 4, October 1962, pp. 75–82.

[3] Sage, A. P., *Optimum Systems Control*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968.

[4] Bellman, R. E., and R. E. Kalaba, *Quasilinearization and Nonlinear Boundary Value Problems*, Elsevier, New York, 1965.

PART II

BASIC THEORY OF THE OPTIMAL REGULATOR

CHAPTER 2

THE STANDARD REGULATOR PROBLEM—I

2.1 A REVIEW OF THE REGULATOR PROBLEM

We shall be concerned almost exclusively with linear finite-dimensional systems, which frequently will also be time invariant. The systems may be represented by equations of the type

$$\dot{x}(t) = F(t)x(t) + G(t)u(t) \tag{2.1-1}$$

$$y(t) = H'(t)x(t). \tag{2.1-2}$$

Here, $F(t)$, $G(t)$, and $H(t)$ are matrix functions of time, in general with continuous entries. If their dimensions are respectively $n \times n$, $n \times m$, $n \times p$, the n vector $x(t)$ denotes the *system state* at time t, the m vector $u(t)$ the *system input* or *system control* at time t, and the p vector $y(t)$ the *system output* at time t.

In classical control work, usually systems with only one input and output are considered. With these restrictions in (2.1-1) and (2.1-2), the vectors $u(t)$ and $y(t)$ become scalars, and the matrices $G(t)$ and $H(t)$ become vectors, and accordingly will often be denoted by lowercase letters to distinguish their specifically vector character. The systems considered are normally also *time invariant*, or *stationary*. In terms of (2.1-1) and (2.1-2), this means that the input $u(t)$ and output $y(t)$ for an initially zero state are related by a time-

invariant impulse response. Furthermore, the most common state-space descriptions of time-invariant systems are those where $F(t)$, $g(t)$, and $h(t)$ are constant with time. (Note, though, that nonconstant $F(t)$, $g(t)$, and $h(t)$ may still define a time-invariant impulse response—e.g., $F(t) = 0$, $g(t) = e^t$, $h(t) = e^{-t}$ defines the time-invariant impulse response $e^{-(t-\tau)}$.)

The classical description of a system is normally in terms of its transfer function matrix, which we denote by $W(s)$, s being the Laplace transform variable. The well-known connection between $W(s)$ and the matrices of (2.1-1) and (2.1-2), if these are constant, is

$$W(s) = H'(sI - F)^{-1}G. \tag{2.1-3}$$

A common class of control problems involves a *plant*, for which a control is desired to achieve one of the following aims:

1. *Qualitative statement of the regulator problem.* Suppose that initially the plant output, or any of its derivatives, is nonzero. Provide a plant input to bring the output and its derivatives to zero. In other words, the problem is to apply a control to take the plant from a nonzero state to the zero state. This problem may typically occur where the plant is subjected to unwanted disturbances that perturb its output (e.g., a radar antenna control system with the antenna subject to wind gusts).
2. *Qualitative statement of the tracking (or servomechanism) problem.* Suppose that the plant output, or a derivative, is required to track some prescribed function. Provide a plant input that will cause this tracking (e.g., when a radar antenna is to track an aircraft such a control is required).

In a subsequent chapter, we shall discuss the tracking problem. For the moment, we restrict attention to the more fundamental regulator problem.

When considering the regulator problem using classical control theory, we frequently seek a solution that uses *feedback* of the output and its derivatives to generate a control. A *controller*, describable by a transfer function, is interposed between the plant output and plant input, with the plant output constituting the controller input. The controller output is the plant input. In Fig. 2.1-1, the feedback arrangement is shown. Both the plant and controller have a single input and output, and are time invariant—i.e., possess a transfer function.

In the modern approach, it is often assumed that the plant states are available for measurement. (If this is not the case, it is generally possible to construct a physical device called a *state estimator*, which produces at its output the plant states, when driven by both the plant input and output. This will be discussed in a later chapter.) In addition to assuming availability of the states, the design of controllers is often restricted by a requirement

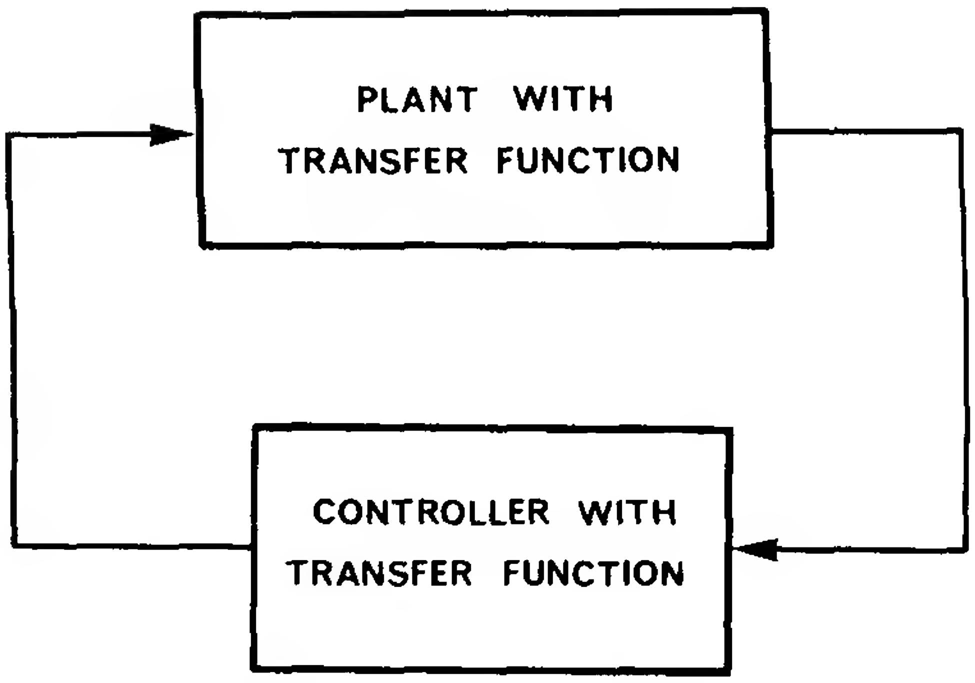


Fig. 2.1-1 Classical feedback arrangement.

that they be nondynamic, or memoryless. In other words, the controller output or plant input $u(t)$ is required to be an instantaneous function of the plant state $x(t)$. The nature of this function may be permitted to vary with time, in which case we could write down a *control law*

$$u(t) = k(x(t), t) \tag{2.1-4}$$

to indicate the dependence of $u(t)$ on both $x(t)$ and t.

Of interest from the viewpoint of ease of implementation is the case of the *linear control law*, given by

$$u(t) = K'(t)x(t) \tag{2.1-5}$$

for some matrix K of appropriate dimension. [When $K(t)$ is a constant matrix, (2.1-5) becomes a constant or time-invariant control law, and, as will become clear, a number of connections with the classical approach can be made.] Figure 2.1-2 shows the modern control arrangement. The plant is assumed to be linear, but may have multiple inputs and outputs and may be time varying. The state estimator constructs the plant state vector from the

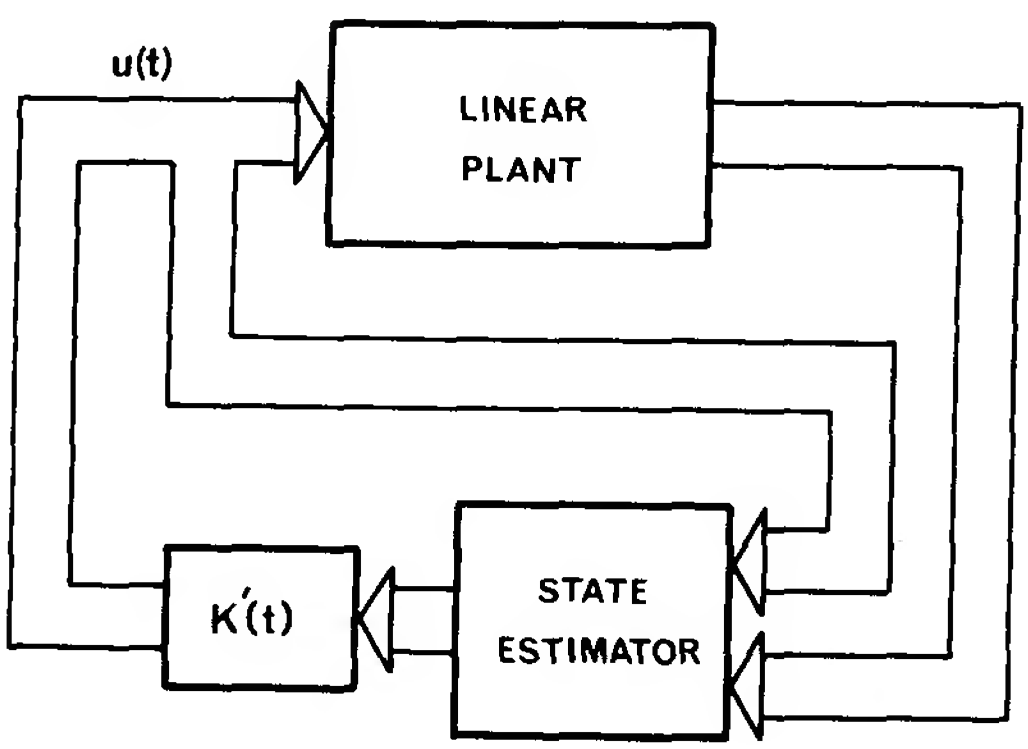


Fig. 2.1-2 Modern control feedback arrangement.

input and output vectors, and is actually a linear, finite-dimensional system itself. Linear combinations of the states are fed back to the system input in accordance with (2.1-5).

When attempting to construct a controller for the regulator problem, we might imagine that the way to proceed would be to search for a control scheme that would take an arbitrary nonzero initial state to the zero state, preferably as fast as possible. Could this be done? If F and G are constant, and if the pair $[F, G]$ is *completely controllable*, the answer is certainly yes [1]. Recall (see Appendix B), that the definition of complete controllability requires that there be a control taking any nonzero state $x(t_0)$ at time t_0 to the zero state at some time T. In fact, if F and G are constant, T can be taken as close to t_0 as desired, and for some classes of time-varying $F(t)$ and $G(t)$, this is also the case. What, therefore, would be wrong with such a scheme? Two things. First, the closer T is to t_0 the greater is the amount of control energy (and the greater is the magnitude of the control) required to effect the state transfer. In any engineering system, an upper bound is set on the magnitude of the various variables in the system by practical considerations. Therefore, one could not take T arbitrarily close to t_0 without exceeding these bounds. Second, as reference to [1] will show, the actual control cannot be implemented as a linear feedback law, for finite T, unless one is prepared to tolerate infinite entries in $K(T)$—i.e., the controller gain at time T. In effect, a linear feedback control is ruled out.

Any other control scheme for which one or both of these objections is valid is equally unacceptable. In an effort to meet the first objection, one could conceive that it is necessary to keep some measure of control magnitude bounded or even small during the course of a control action; such measures might be

$$\int_{t_0}^{T} u'(t)u(t)\,dt, \quad \int_{t_0}^{T} [u'(t)u(t)]^{1/2}\,dt, \quad \max_{t\in[t_0,T]} \|u(t)\|,$$
$$\text{or} \quad \int_{t_0}^{T} u'(t)R(t)u(t)\,dt$$

where the superscript prime denotes matrix (or vector) transposition, and $R(t)$ is a positive definite matrix for all t, which, without loss of generality, can always be taken as symmetric. We shall discuss subsequently how to meet the second objection. Meanwhile, we shall make further adjustments to our original aim of regulation.

First, in recognition of the fact that "near enough is often good enough" for engineering purposes, we shall relax the aim that the system should actually achieve the zero state, and require merely that the state as measured by some norm should become small. If there were some fixed time T by which this was required, we might ask that $x'(T)Ax(T)$, with A some positive definite matrix, be made small.

Second, it is clearly helpful from the control point of view to have $||x(t)||$ small for any t in the interval over which control is being exercized, and we can express this fact by asking (for example) that $\int_{t_0}^T x'(t)Q(t)x(t)\,dt$ be small, where $Q(t)$ is symmetric positive definite. In some situations, as we shall see, it proves sufficient to have $Q(t)$ nonnegative definite.

Let us now sum up the desirable properties of a regulator system.

1. The regulator system should involve a linear control law, of the form $u(t) = K'(t)x(t)$.
2. The regulator scheme should ensure the smallness of quantities such as $\int_{t_0}^T u'(t)R(t)u(t)\,dt$, $x'(T)Ax(T)$, and $\int_{t_0}^T x'(t)Q(t)x(t)\,dt$, where $R(\cdot)$, A, and $Q(\cdot)$ have the positivity properties mentioned earlier.

For the moment, we shall concentrate on 2. Subsequently, we shall see that a regulator scheme with property 2 automatically possesses property 1 too.

To consider achieving property 2, let us define the *performance index*

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^T (u'Ru + x'Qx)\,dt + x'(T)Ax(T). \tag{2.1-6}$$

As the notation implies, the value taken by V depends on the initial state $x(t_0)$ and time t_0, and the control over the interval $[t_0, T]$.

A reasonable approach to achieving property 2 would be to choose that control $u(\cdot)$ which minimized the performance index (2.1-6) [always assuming such a $u(\cdot)$ exists]. This idea leads to the following formal statement of the regulator problem.

Regulator problem. Consider the system (2.1-1), where the entries of $F(t)$ and $G(t)$ are assumed to be continuous. Let the matrices $Q(t)$ and $R(t)$ have continuous entries, be symmetric, and be nonnegative and positive definite, respectively. Let A be a nonnegative definite symmetric matrix. Define the *performance index* $V(x(t_0), u(\cdot), t_0)$ as in (2.1-6) and the *minimization problem* as the task of finding an *optimal control* $u^*(t)$, $t \in [t_0, T]$, minimizing V, and the associated *optimum performance index* $V^*(x(t_0), t_0)$—i.e., the value of V obtained by using the optimal control.

Notice that earlier it was suggested that A should be positive definite, whereas the statement of the regulator problem merely suggests that it should be nonnegative definite. As we shall see subsequently, the size of the final state $x(T)$ can frequently be made small merely by the relaxed requirement. Indeed, the choice $A = 0$ will often lead to a satisfactory result.

As already remarked, the minimization of (2.1-6) turns out to be achievable with a linear feedback law. This is the reason why the performance index includes a measure of control energy in the form $\int_0^T u'(t)R(t)u(t)\,dt$,

rather than, say, $\int_0^T [u'(t)u(t)]^5\,dt$, or any of the other measures suggested. These other measures do not, in general, lead to linear feedback laws. The form of the other terms in the performance index (2.1-6) is also generally necessary to obtain linearity of the feedback law.

Before studying the minimization problem further, we note the following references. Books [2] and [3] are but two of a number of excellent treatments of the regulator problem, and, of course, optimal control in general. Several papers dealing with the regulator problem could be read with benefit [4–7].

Problem 2.1-1. Consider the system

$$\dot{x} = F(t)x + G(t)u$$

with F, G possessing continuous entries. Show that there does not exist a control law

$$u = K'(t)x(t)$$

with the entries of $K(t)$ continuous, such that with arbitrary $x(t_0)$ and some finite T, $x(T) = 0$. [*Hint:* Use the fact that if $\dot{x} = \bar{F}(t)x$ where $\bar{F}$ has continuous entries then a transition matrix exists.]

Problem 2.1-2. Electrical networks composed of a finite number of interconnected resistors, capacitors, inductors, and transformers can normally be described by state-space equations of the form

$$\dot{x} = Fx + Gu$$
$$y = H'x + Ju.$$

The entries of the state vector will often correspond to capacitor voltages and inductor currents, the entries of the input vector to the currents at the various ports of the network, and the entries of the output vector to the voltages at the various ports of the network. Assuming the initial $x(t_0)$ is nonzero, give a physical interpretation to the problem of minimizing

$$\int_{t_0}^{T} (u'Ju + x'Hu)\,dt.$$

2.2 THE HAMILTON-JACOBI EQUATION

In this section, we temporarily move away from the specific regulator problem posed in the last section to consider a wider class of optimization problems requiring the minimization of a performance index. We shall, in fact, derive a partial differential equation, *the Hamilton–Jacobi equation*, satisfied by the optimal performance index under certain differentiability and continuity assumptions. Moreover, it can be shown that if a solution to the Hamilton–Jacobi equation has certain differentiability properties, then this solution is the desired performance index. But since such a solution

need not exist, and not every optimal performance index satisfies the Hamilton–Jacobi equation, the equation only represents a sufficient, rather than a necessary, condition on the optimal performance index.

In this section, we shall also show how the optimal performance index, if it satisfies the Hamilton–Jacobi equation, determines an optimal control. This will allow us in Sec. 2.3 to combine the statements of the regulator problem and the Hamilton–Jacobi theory, to deduce the optimal performance index and associated optimal control for the regulator problem.

Other approaches will lead to the derivation of the optimal control and optimal performance index associated with the regulator problem, notably the use of the Maximum Principle of Pontryagin, combined with the Euler–Lagrange equations, [2], [3], and [8]. The Maximum Principle and Euler–Lagrange equations are lengthy to derive, although their application to the regulator problem is straightforward. The simplest route to take without quoting results from elsewhere appears to be the development of the Hamilton–Jacobi equation with subsequent application to the regulator problem. Actually, the Hamilton–Jacobi equation has so far rarely proved useful except for linear regulator problems, to which it seems particularly well suited.

The treatment we follow in deducing the Hamilton–Jacobi equation is a blend of treatments to be found in [3] and [7]. We start by posing the following optimal control problem. For the system

$$\dot{x} = f(x, u, t) \qquad x(t_0) \text{ given} \tag{2.2-1}$$

find the optimal control $u^*(t)$, $t \in [t_0, T]$, which minimizes

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)). \tag{2.2-2}$$

Without explicitly defining for the moment the degree of smoothness—i.e., the number of times quantities should be differentiable—we shall restrict f, l, and m to being smooth functions of their arguments. Otherwise, $f(x, u, t)$ can be essentially arbitrary, whereas $l(x(\tau), u(\tau), \tau)$ and $m(x(T))$ will often be nonnegative, to reflect some physical quantity the minimization of which is desired. As the notation implies, the performance index depends on the initial state $x(t_0)$ and time t_0, and the control $u(t)$ for all $t \in [t_0, T]$. The optimal control $u^*(\cdot)$ may be required a priori to lie in some special set, such as the set of piecewise continuous functions, square-integrable functions bounded by unity, etc.

Let us adopt the notation $u_{[a, b]}$ to denote a function $u(\cdot)$ restricted to the interval $[a, b]$. Let us also make the definition

$$V^*(x(t), t) = \min_{u[t, T]} V(x(t), u(\cdot), t). \tag{2.2-3}$$

That is, if the system starts in state $x(t)$ at time t, the minimum value of the performance index (2.2-2) is $V^*(x(t), t)$. Notice that $V^*(x(t), t)$ is independent

of $u(\cdot)$, precisely because knowledge of the initial state and time abstractly determines the particular control, by the requirement that the control minimize $V(x(t), u(\cdot), t)$. Rather than just searching for the control minimizing (2.2-2) and for the value of $V^*(x(t_0), t_0)$ for various $x(t_0)$, we shall study the evaluation of (2.2-3) for all t and $x(t)$, and the determination of the associated optimum control. Of course, assuming we have a functional expression for V^* in terms of $x(t)$ and t, together with the optimal control, we solve the optimization problem defined by (2.2-1) and (2.2-2) by setting $t = t_0$.

Now, for arbitrary t in the range $[t_0, T]$ and t_1 in the range $[t, T]$, we have

$$\begin{aligned}
V^*(x(t), t) &= \min_{u[t,T]}\left[\int_t^T l(x(\tau), u(\tau), \tau)\,d\tau + m(x(T))\right] \\
&= \min_{u[t,t_1)}\left\{\min_{u[t_1,T]}\left[\int_t^{t_1} l(x(\tau), u(\tau), \tau)\,d\tau \right.\right. \\
&\qquad \left.\left. + \int_{t_1}^T l(x(\tau), u(\tau), \tau)\,d\tau + m(x(T))\right]\right\} \\
&= \min_{u[t,t_1)}\left\{\int_t^{t_1} l(x(\tau), u(\tau), \tau)\,d\tau \right. \\
&\qquad \left. + \min_{u[t_1,T]}\left[\int_{t_1}^T l(x(\tau), u(\tau), \tau)\,d\tau + m(x(T))\right]\right\},
\end{aligned}$$

or

$$V^*(x(t), t) = \min_{u[t,t_1)}\left[\int_t^{t_1} l(x(\tau), u(\tau), \tau)\,d\tau + V^*(x(t_1), t_1)\right]. \tag{2.2-4}$$

In (2.2-4), let us now set $t_1 = t + \Delta t$, where Δt is small. Applying Taylor's theorem to expand the right-hand side (noting that the smoothness assumptions permit us to do this), we obtain

$$\begin{aligned}
V^*(x(t), t) = \min_{u[t,t+\Delta t)} &\left\{\Delta t\, l(x(t + \alpha\,\Delta t), u(t + \alpha\,\Delta t), t + \alpha\,\Delta t)\right. \\
&+ V^*(x(t), t) + \left[\frac{\partial V^*}{\partial x}(x(t), t)\right]' \frac{dx(t)}{dt}\Delta t \\
&\left. + \frac{\partial V^*}{\partial t}(x(t), t)\,\Delta t + 0\,(\Delta t)^2\right\}
\end{aligned}$$

where α is some constant lying between 0 and 1. Immediately,

$$\begin{aligned}
\frac{\partial V^*}{\partial t}(x(t), t) = -\min_{u[t,t+\Delta t)} &\left\{l(x(t + \alpha\,\Delta t), u(t + \alpha\,\Delta t), t + \alpha\,\Delta t)\right. \\
&\left. + \left[\frac{\partial V^*}{\partial x}(x(t), t)\right]' f(x(t), u(t), t) + 0\,(\Delta t)\right\}.
\end{aligned}$$

Now, let Δt approach zero, to conclude that

$$\frac{\partial V^*}{\partial t}(x(t), t) = -\min_{u(t)}\left\{l(x(t), u(t), t) + \left[\frac{\partial V^*}{\partial x}(x(t), t)\right]' f(x(t), u(t), t)\right\}.$$

In this equation, f and l are known functions of their arguments, whereas V^* is unknown. In order to emphasize this point, we shall rewrite the equation as

$$\frac{\partial V^*}{\partial t} = -\min_{u(t)}\left[l(x(t), u(t), t) + \frac{\partial V^{*\prime}}{\partial x} f(x(t), u(t), t)\right]. \tag{2.2-5}$$

This is one statement of the Hamilton–Jacobi equation. In this format, it is not precisely a partial differential equation but a mixture of a functional and a partial differential equation.

The value of $u(t)$ minimizing the right-hand side of (2.2-5) will depend on the values taken by $x(t)$, $\partial V^*/\partial x$, and t. In other words, the minimizing $u(t)$ is an instantaneous function—call it $\bar{u}(x(t), \partial V^*/\partial x, t)$—of the three variables $x(t)$, $\partial V^*/\partial x$, t. [To be sure, if the *explicit* way $\partial V^*/\partial x$ depended on $x(t)$ and t were known, at this point one could express the minimizing $u(t)$ as a function of merely $x(t)$ and t. But at this point, the explicit functions that V^*, and $\partial V^*/\partial x$, are of $x(t)$ and t have yet to be found.]

The preceding derivations also tell us that to minimize $V(x(t), u(\cdot), t)$, the value of the minimizing control at time t is precisely $\bar{u}(x(t), \partial V^*/\partial x, t)$.

With the above definition of $\bar{u}(\cdot, \cdot, \cdot)$, Eq. (2.2-5) becomes

$$\begin{aligned}\frac{\partial V^*}{\partial t} &= -l[x(t), \bar{u}(x(t), \frac{\partial V^*}{\partial x}, t), t] \\ &\quad - \frac{\partial V^{*\prime}}{\partial x} f[(x(t), \bar{u}(x(t), \frac{\partial V^*}{\partial x}, t), t].\end{aligned} \tag{2.2-6}$$

Despite the bewildering array of symbols within symbols, (2.2-6) is but a first-order partial differential equation with one dependent variable, V^*, and two independent variables, $x(t)$ and t, because l, f, and $\bar{u}$ are known functions of their arguments.

A boundary condition for (2.2-6) is very simply derived. Reference to the performance index (2.2-2) shows that $V(x(T), u(\cdot), T) = m(x(T))$ for all $u(\cdot)$, and, accordingly, the minimum value of this performance index with respect to $u(\cdot)$ is also $m(x(T))$. That is,

$$V^*(x(T), T) = m(x(T)). \tag{2.2-7}$$

The pair (2.2-6) and (2.2-7) may also be referred to as the Hamilton–Jacobi equation, and constitute a true partial differential equation.

If the minimization implied by (2.2-5) is impossible—i.e., if $\bar{u}(\cdot, \cdot, \cdot)$ does not exist—then the whole procedure is invalidated and the Hamilton–Jacobi approach cannot be used in tackling the optimization problem.

We now consider how to determine the optimal control for the problem defined by (2.2-1) and (2.2-2). We assume that (2.2-6) and (2.2-7) have been solved so that $V^*(x(t), t)$ is a known function of $x(t)$ and t. Now, $V^*(x(t), t)$ determines a function of $x(t)$ and t—call it $\hat{u}(x(t), t)$—via the formula

$$\hat{u}(x(t), t) = \bar{u}[x(t), \frac{\partial V^*}{\partial x}(x(t), t), t]. \tag{2.2-8}$$

That is, $\hat{u}$ is the same as $\bar{u}$, except that the second variable on which $\bar{u}$ depends itself becomes a specified function of the first and third variables.

This new function $\hat{u}(\cdot, \cdot)$ has two important properties. The first and more easily seen is that $\hat{u}(x(t), t)$ is the value at time t of the optimal control minimizing

$$V(x(t), u(\cdot), t) = \int_t^T l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)). \tag{2.2-9}$$

That is, to achieve the optimal performance index $V^*(x(t), t)$, we start off with a control $\hat{u}(x(t), t)$. That this is so is implicit in the arguments leading to (2.2-5), as previously shown.

The second property is that the optimal control $u^*(\cdot)$ for the *original* minimization problem defined by (2.2-1) and (2.2-2)—with t_0 as the initial time and t as an intermediate value of time—is related to $\hat{u}(\cdot, \cdot)$ simply by

$$u^*(t) = \hat{u}(x(t), t) \tag{2.2-10}$$

when $x(t)$ is the state at time t arising from application of $u^*(\cdot)$ over $[t_0, t)$.

To some, this result will be intuitively clear. To demonstrate it, we examine a variant of the arguments leading to (2.2-4). By definition,

$$V^*(x(t_0), t_0) = \min_{u[t_0,T]} \left[\int_{t_0}^T l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)) \right],$$

and the minimum is achieved by $u^*(\cdot)$. With $u^*(\cdot)$ regarded as the sequential use of $u^*_{[t_0,t)}$ and $u^*_{[t,T]}$, and *with the assumption that* $u_{[t_0,t)}$ *is applied until time* t, evidently

$$V^*(x(t_0), t_0) = \min_{u[t,T]} \left[\int_{t_0}^T l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)) \right] \tag{2.2-11}$$

$$= \int_{t_0}^t l(x(\tau), u(\tau), \tau)\, d\tau$$

$$+ \min_{u[t,T]} \left[\int_t^T l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)) \right]. \tag{2.2-12}$$

The minimization in (2.2-11), and therefore (2.2-12), is achieved by $u^*_{[t,T]}$. In other words, $u^*_{[t,T]}$ is the optimal control for the system (2.2-1) with performance index

$$\int_t^T l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)) \tag{2.2-13}$$

with initial state $x(t)$, where $x(t)$ is derived by starting (2.2-1) at time t_0 in state $x(t_0)$, and applying $u^*_{[t_0,t)}$. (This is, in fact, one statement of the Principle of Optimality, [10], to the effect that a control policy optimal over an interval $[t_0, T]$ is optimal over all subintervals $[t, T]$.) But $\hat{u}(x(t), t)$ is the value of the optimal control at time t for the performance index (2.2-13), and so

$$\hat{u}(x(t), t) = u^*_{[t,T]}(t) = u^*(t). \tag{2.2-14}$$

Several points should now be noted. First, because of the way it is calculated $\hat{u}(x(t), t)$ is independent of t_0. The implication is, then, that the optimal control at an arbitrary time t for the minimization of

$$V(x(\sigma), u(\cdot), \sigma) = \int_{\sigma}^{T} l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)) \qquad (2.2\text{-}15)$$

is also $\hat{u}(x(t), t)$. Put another way, the control $\hat{u}(x(\cdot), \cdot)$ is the optimal control for the whole class of problems (2.2-15), with variable $x(\sigma)$ and σ.

The second point to note is that the optimal control at time t is given in terms of the state $x(t)$ at time t, although, because its functional dependence on the state may not be constant, it is in general a time-variable function of the state. It will be theoretically implementable with a feedback law, as in Fig. 2.2-1. (Other schemes, such as the Maximum Principle and

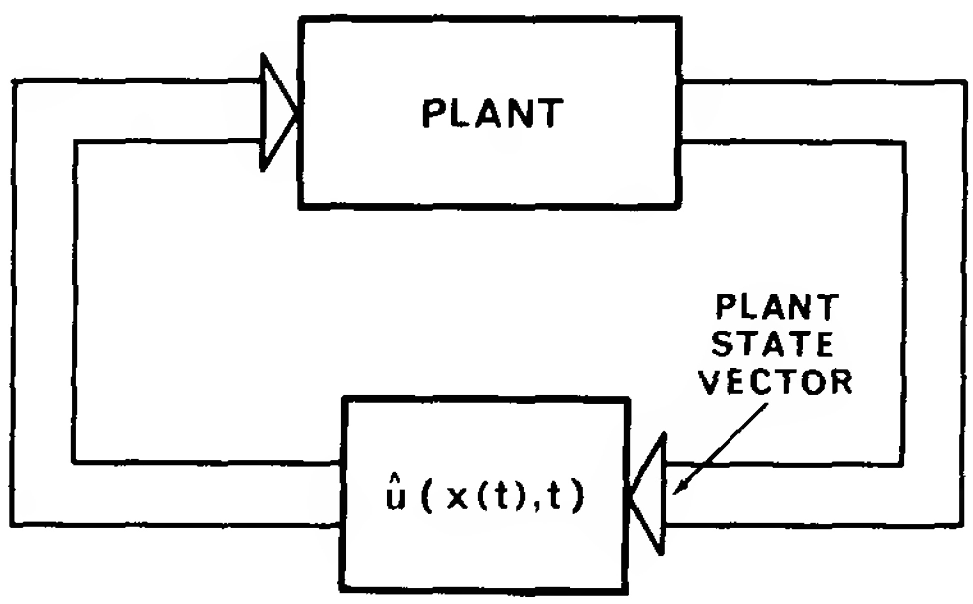


Fig. 2.2-1 Feedback implementation of the optimal control.

Euler–Lagrange equations, for computing the optimal control do not necessarily have this useful property; in these schemes, the optimal control may often be found merely as a certain function of time.)

The third point is that the remarks leading to the Hamilton–Jacobi equation are reversible, in the sense that if a suitably smooth solution of the equation is known, this solution has to be the optimal performance index $V^*(x(t), t)$.

Finally, rigorous arguments, as in, for example, [4] and [5], pin down the various smoothness assumptions precisely and lead to the following conclusion, which we shall adopt as a statement of the Hamilton–Jacobi results.

Hamilton-Jacobi equation. Consider the system

$$\dot{x} = f(x, u, t) \qquad (2.2\text{-}1)$$

and the performance index

$$V(x(t), u(\cdot), t) = \int_{t}^{T} l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)). \qquad (2.2\text{-}9)$$

Suppose f, l, and m are continuously differentiable in all arguments, that there exists an absolute minimum† of $l(x, u, t) + \lambda' f(x, u, t)$ with respect to $u(t)$ of the form $\bar{u}(x(t), \lambda, t)$, and that $\bar{u}$ is continuously differentiable in all its arguments. Furthermore, suppose that $V^*(x(t), t)$ is a solution of the Hamilton–Jacobi equation (2.2-5) or (2.2-6) with boundary condition (2.2-7). Then $V^*(\cdot, \cdot)$ is the optimal performance index for (2.2-9), and the control given by (2.2-8) is the optimal control at time t for the class of problems with performance index

$$V(x(\sigma), u(\cdot), \sigma) = \int_\sigma^T l(x(\tau), u(\tau), \tau)\, d\tau + m(x(T)). \tag{2.2-15}$$

Conversely, suppose that f, l, and m are continuously differentiable in all arguments, that there exists an optimal control, and that the corresponding minimum value of (2.2-9), $V^*(x(t), t)$, is twice continuously differentiable in its arguments. Suppose also that $l(x, u, t) + (\partial V^*/\partial x)' f(x, u, t)$ has an absolute minimum with respect to $u(t)$ at $\hat{u}(x(t), t)$, and that $u(\cdot, \cdot)$ is differentiable in x and continuous in t. Then $V^*(\cdot, \cdot)$ satisfies the Hamilton–Jacobi equation (2.2-5) or (2.2-6) with boundary condition (2.2-7).

We conclude this section with a simple example illustrating the derivation of the Hamilton–Jacobi equation in a particular instance. Suppose we are given the system

$$\dot{x} = u,$$

with performance index

$$V(x(0), u(\cdot), 0) = \int_0^T (u^2 + x^2 + \tfrac{1}{2}x^4)\, dt.$$

Using Eq. (2.2-5), we have

$$\frac{\partial V^*}{\partial t} = -\min_{u(t)} \left\{u^2 + x^2 + \tfrac{1}{2}x^4 + \frac{\partial V^*}{\partial x} u\right\}.$$

The minimizing $u(t)$ is clearly

$$\bar{u} = -\frac{1}{2}\frac{\partial V^*}{\partial x},$$

and we have

$$\frac{\partial V^*}{\partial t} = \frac{1}{4}\left(\frac{\partial V^*}{\partial x}\right)^2 - x^2 - \frac{1}{2}x^4$$

as the Hamilton–Jacobi equation for this problem, with boundary condition $V^*(x(T), T) = 0$. The question of how this equation might be solved is

†Though the point is inessential to our development, we should note that $u(t)$ may be constrained a priori to lie in some set $U(t)$ strictly contained in the Euclidean space of dimension equal to the dimension of u. All minimizations are then subject to the constraint $u(t) \in U(t)$.

quite unresolved by the theory presented so far. In actual fact, it is rarely possible to solve a Hamilton–Jacobi equation, although for the preceding example, a solution happens to be available [11]. It is extraordinarily complex, and its repetition here would serve no purpose.

Problem 2.2-1. Consider a system of the form

$$\dot{x} = f(x) + gu$$

with performance index

$$V(x(t), u(\cdot), t) = \int_t^T (u^2 + g(x))\, dt.$$

Show that the Hamilton–Jacobi equation is linear in $\partial V^*/\partial t$ and quadratic in $\partial V^*/\partial x$.

Problem 2.2-2. Let u be an r vector, and let λ and x be n vectors. Let A, B, C be constant matrices of appropriate dimensions such that the following function of u, x, and p can be formed:

$$Q(u, x, p) = u'Au + 2x'Bu + 2u'Cp.$$

Show Q has a unique minimum in u for all x and p if, and only if, $\frac{1}{2}(A + A')$ is positive definite.

2.3 SOLUTION OF THE FINITE-TIME REGULATOR PROBLEM

In this section, we return to the solution of the regulator problem, which we restate for convenience.

Regulator problem. Consider the system

$$\dot{x} = F(t)x(t) + G(t)u(t) \qquad x(t_0) \text{ given} \tag{2.3-1}$$

with the entries of $F(t)$, $G(t)$ assumed continuous. Let the matrices $Q(t)$ and $R(t)$ have continuous entries, be symmetric, and be nonnegative and positive definite, respectively. Let A be a nonnegative definite matrix. Define the *performance index*,

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^T (u'Ru + x'Qx)\, dt + x'(T)Ax(T), \tag{2.3-2}$$

and the *minimization problem* as the task of finding an *optimal control* $u^*(t)$, $t\epsilon[t_0, T]$, minimizing V and the associated *optimum performance index* $V^*(x(t_0), t_0)$.

For the moment, assume that T is finite.

To solve the problem, we shall make use of the results on the Hamilton–Jacobi equation summarized at the end of the last section. An outline of the problem solution follows.

1. We shall show by simple arguments independent of the Hamilton–Jacobi theory that the optimal performance index $V^*(x(t), t)$, if it exists, must be of the form $x'(t)P(t)x(t)$, where $P(t)$ is a symmetric matrix.
2. With the assumption that $V^*(x(t), t)$ exists, result 1 will be used together with the Hamilton–Jacobi theory to show that $P(t)$ satisfies a nonlinear differential equation—in fact, a matrix Riccati equation.
3. We shall establish existence of $V^*(x(t), t)$.
4. We shall find the optimal control.

To carry out this program, it is necessary to make the following temporary assumption.

TEMPORARY ASSUMPTION 2.3-1. Assume that $F(t)$, $G(t)$, $R(t)$, and $Q(t)$ have entries that are continuously differentiable.

This assumption is removed in Prob. 2.3-2.

We note that some treatments of the regulator problem assume a priori the form $x'(t)P(t)x(t)$ for the optimal performance index. It is therefore interesting to observe a simple derivation of this form appearing in [12].

The quadratic form of V*(x(t), t). The necessary and sufficient conditions for $V^*(x(t), t)$ to be a quadratic form are that

$$V^*(\lambda x, t) = \lambda^2 V^*(x, t) \qquad \text{for all real } \lambda \tag{2.3-3}$$

$$V^*(x_1, t) + V^*(x_2, t) = \tfrac{1}{2}[V^*(x_1 + x_2, t) + V^*(x_1 - x_2, t)]. \tag{2.3-4}$$

[The student is asked to verify this claim in Prob. (2.3-1).] To show that (2.3-3) and (2.3-4) hold, we adopt the temporary notation u_x^* to denote the optimal control over $[t, T]$ when the initial state is $x(t)$ at time t. Then the linearity of (2.3-1) and the quadratic nature of (2.3-2) imply the following equalities, whereas the inequalities follow directly from the fact that an optimal index is the minimum index. We have that

$$V^*(\lambda x, t) \leq V(\lambda x, \lambda u_x^*(\cdot), t) = \lambda^2 V^*(x, t)$$
$$\lambda^2 V^*(x, t) \leq \lambda^2 V(x, \lambda^{-1} u_{\lambda x}^*(\cdot), t) = V^*(\lambda x, t)$$

for all real constants λ. These imply (2.3-3) directly. Similar reasoning gives the inequality

$$\begin{aligned} V^*(x_1, t) + V^*(x_2, t) &= \tfrac{1}{4}[V^*(2x_1, t) + V^*(2x_2, t)] \\ &\leq \tfrac{1}{4}[V(2x_1, u_{x_1+x_2}^* + u_{x_1-x_2}^*, t) \\ &\qquad + V(2x_2, u_{x_1+x_2}^* - u_{x_1-x_2}^*, t)] \\ &= \tfrac{1}{2}[V(x_1 + x_2, u_{x_1+x_2}^*, t) \\ &\qquad + V(x_1 - x_2, u_{x_1-x_2}^*, t)] \\ &= \tfrac{1}{2}[V^*(x_1 + x_2, t) + V^*(x_1 - x_2, t)]. \end{aligned} \tag{2.3-5}$$

By making use of the controls $u^*_{x_1}$ and $u^*_{x_2}$, we establish the following inequality in a like manner:

$$\tfrac{1}{2}[V^*(x_1 + x_2, t) + V^*(x_1 - x_2, t)] \leq V^*(x_1, t) + V^*(x_2, t). \tag{2.3-6}$$

Then (2.3-5) and (2.3-6) imply (2.3-4).

We conclude that $V^*(x(t), t)$ has the form

$$V^*(x(t), t) = x'(t)P(t)x(t) \tag{2.3-7}$$

for some matrix $P(t)$, without loss of generality symmetric. {If $P(t)$ is not symmetric, it may be replaced by the symmetric matrix $\frac{1}{2}[P(t) + P'(t)]$ without altering (2.3-7)}.

Derivation of the matrix Riccati equation. Now we shall show, using the Hamilton–Jacobi equation, that the symmetric matrix $P(t)$ satisfies a matrix Riccati equation.

The first form of the Hamilton–Jacobi equation is now repeated:

$$\frac{\partial V^*(x(t), t)}{\partial t} = -\min_{u(t)} \left\{ l(x(t), u(t), t) + \left[\frac{\partial V^*}{\partial x}(x(t), t)\right]' f(x(t), u(t), t) \right\}. \tag{2.3-8}$$

In our case, $l(x(t), u(t), t)$ is $u'(t)R(t)u(t) + x'(t)Q(t)x(t)$; $[(\partial V^*/\partial x)(x(t), t)]'$ from (2.3-7) is $2x'(t)P(t)$, whereas $f(x(t), u(t), t)$ is simply $F(t)x(t) + G(t)u(t)$. The left side of (2.3-8) is simply $x'(t)\dot{P}(t)x(t)$. Hence, Eq. (2.3-8) becomes, in the special case of the regulator problem,

$$x'\dot{P}x = -\min_{u(t)} [u'Ru + x'Qx + 2x'PFx + 2x'PGu]. \tag{2.3-9}$$

To find the minimum of the expression on the right-hand side of (2.3-9), we simply note the following identity, obtained by completing the square:

$$\begin{aligned} u'Ru + x'Qx + 2x'PFx + 2x'PGu = {} & (u + R^{-1}G'Px)'R(u + R^{-1}G'Px) \\ & + x'(Q - PGR^{-1}G'P \\ & + PF + F'P)x. \end{aligned}$$

Because the matrix $R(t)$ is positive definite, it follows that (2.3-9) is minimized by setting

$$\bar{u}(t) = -R^{-1}(t)G'(t)P(t)x(t), \tag{2.3-10}$$

in which case one obtains

$$\begin{aligned} x'(t)\dot{P}(t)x(t) = {} & -x'(t)[Q(t) - P(t)G(t)R^{-1}(t)G'(t)P(t) \\ & + P(t)F(t) + F'(t)P(t)]x(t). \end{aligned}$$

Now this equation holds for all $x(t)$; therefore,

$$-\dot{P}(t) = P(t)F(t) + F'(t)P(t) - P(t)G(t)R^{-1}(t)G'(t)P(t) + Q(t) \tag{2.3-11}$$

where we use the fact that both sides are symmetric.

Equation (2.3-11) is the matrix Riccati equation we are seeking. It has a boundary condition following immediately from the Hamilton–Jacobi boundary condition. We recall that $V^*(x(T), T) = m(x(T))$, which implies in the regulator problem that $x'(T)P(T)x(T) = x'(T)Ax(T)$. Since both $P(T)$ and A are symmetric, and $x(T)$ is arbitrary,

$$P(T) = A. \tag{2.3-12}$$

Before we proceed further, it is proper to examine the validity of the preceding manipulations in the light of the statement of the Hamilton–Jacobi equation at the end of the last section. Observe the following.

1. The minimization required in (2.3-9) is, in fact, possible, yielding the continuously differentiable minimum of (2.3-10). [In the notation of the last section, the expression on the right of (2.3-10) is the function $\bar{u}(\cdot, \cdot, \cdot)$ of $x(t)$, $\partial V^*/\partial x$, and t.]
2. The loss function $x'Qx + u'Ru$ and function $Fx + Gu$ appearing in the basic system equation have the necessary differentiability properties, this being guaranteed by Temporary Assumption 2.3-1.
3. If $P(t)$, the solution of (2.3-11), exists, both $\dot{P}(t)$ and $\ddot{P}(t)$ exist and are continuous, the former because of the relation (2.3-11), the latter because differentiation of both sides of (2.3-11) leads to $\ddot{P}(t)$ being equal to a matrix with continuous entries (again, Temporary Assumption 2.3-1 is required). Consequently, $x'(t)P(t)x(t)$ is twice continuously differentiable.

Noting that Eqs. (2.3-11) and (2.3-12) imply the Hamilton–Jacobi equation (2.3-8), with appropriate initial conditions, we can then use the statement of the Hamilton–Jacobi equation of the last section to conclude the following.

1. If the optimal performance index $V^*(x(t), t)$ exists, it is of the form $x'(t)P(t)x(t)$, and $P(t)$ satisfies (2.3-11) and (2.3-12).
2. If there exists a symmetric matrix $P(t)$ satisfying (2.3-11) and (2.3-12), then the optimal performance index $V^*(x(t), t)$ exists, satisfies the Hamilton–Jacobi equation, and is given by $x'(t)P(t)x(t)$.

In theory, $P(t)$, and in particular $P(t_0)$, can be computed† from (2.3-11) and (2.3-12). Thus, aside from the existence question, the problem of finding the optimal performance index is solved.

Existence of the optimal performance index $V^*(x(t), t)$. Here we shall argue that $V^*(x(t), t)$ must exist for all $t \leq T$. Suppose it does not. Then,

†By the word computed, we mean obtainable via numerical computation, generally by using a digital computer. There is no implication that an analytic formula yields $P(t)$.

by the preceding arguments, Eqs. (2.3-11) and (2.3-12) do not have a solution $P(t)$ for all $t \leq T$.

The standard theory of differential equations yields the existence of a solution of (2.3-11) and (2.3-12) in a neighborhood of T. For points sufficiently far distant from T, a solution may not exist, in which case (2.3-11) exhibits the phenomenon of a finite escape time. That is, moving back earlier in time from T, there is a first time T' such that $P(t)$ exists for all t in $(T', T]$, but as t approaches T', some entry or entries of $P(t)$ become unbounded. Then $P(t)$ fails to exist for $t \leq T'$. Moreover, the only way that the solution of (2.3-11) and (2.3-12) can fail to exist away from T is if there is a finite escape time.

Since our assumption that $V^*(x(t), t)$ does not exist for all $t \leq T$ implies that there is a finite escape time, we shall assume existence of a finite escape time $T' < T$ and show that this leads to a contradiction. We have $V^*(x(t), t)$ exists for all t in $(T', T]$, and, in particular, $V^*(x(T' + \epsilon), T' + \epsilon)$ exists for all positive ϵ less than $(T - T')$. Now

$$0 \leq V^*(x(T' + \epsilon), T' + \epsilon) = x'(T' + \epsilon)P(T' + \epsilon)x(T' + \epsilon),$$

the inequality holding because of the nonnegativity of both the integrand for all u and the final value term in (2.3-2). Hence, $P(T' + \epsilon)$ is a nonnegative definite matrix. As ϵ approaches zero, some entry becomes unbounded; without loss of generality, we can conclude that at least one diagonal entry becomes unbounded. If this were not the case, a certain 2×2 principal minor of $P(T' + \epsilon)$ must become negative as ϵ approaches zero, which contradicts the nonnegative definite property of $P(T' + \epsilon)$. Therefore, we suppose that a diagonal element—say, the ith—is unbounded as ϵ approaches zero; let e_i be a vector with zeros for all entries except the ith, where the entry is 1. Then

$$V^*(e_i, T' + \epsilon) = p_{ii}(T' + \epsilon),$$

which approaches infinity as ϵ approaches zero. (Here, p_{ij} denotes the entry in the ith row and jth column of P.)

But the optimal performance index is never greater than the index resulting from using an arbitrary control. In particular, suppose the zero control is applied to the system (2.3-1), and let $\Phi(t, \tau)$ denote the transition matrix.

Starting in state e_i at time $T' + \epsilon$, the state at time τ is $\Phi(\tau, T' + \epsilon)e_i$, and the associated performance index is

$$V(e_i, 0, T' + \epsilon) = \int_{T'+\epsilon}^{T} e_i\Phi'(\tau, T' + \epsilon)Q(\tau)\Phi(\tau, T' + \epsilon)e_i \, d\tau + e_i'\Phi'(T, T' + \epsilon)A\Phi(T, T' + \epsilon)e_i,$$

which must not be smaller than $p_{ii}(T' + \epsilon)$. But as ϵ approaches zero, $V(e_i, 0, T' + \epsilon)$ plainly remains bounded, whereas $p_{ii}(T' + \epsilon)$ approaches

infinity. Hence, we have a contradiction that rules out the existence of a finite escape time for (2.3-11).

Thus, (2.3-11) and (2.3-12) define $P(t)$ for all $t \leq T$, and therefore the index $V^*(x(t), t) = x'(t)P(t)x(t)$ exists for all $t \leq T$.

The optimal control. In the course of deriving the Riccati equation, we found the optimal control at time t for the regulator problem with initial time t when constructing the minimizing $u(t)$ of (2.3-9) in Eq. (2.3-10). But, as pointed out in the last section, this gives the optimal control $u^*(\cdot)$ for an arbitrary initial time [see (2.2-10)]; thus,

$$u^*(t) = -R^{-1}(t)G'(t)P(t)x(t). \tag{2.3-13}$$

Note that in (2.3-10), $P(t)$ was unknown. The product $2P(t)x(t)$ represents $\partial V^*/\partial x$, and $\bar{u}(t)$ was to be regarded as being defined by independent variables $x(t)$ (actually absent from the functional form for $\bar{u}$), $\partial V^*/\partial x$, and t. Subsequently, we were able to express $\partial V^*/\partial x$ explicitly in terms of t and x, since $P(t)$ became explicitly derivable from (2.3-11). This led to the feedback law of (2.3-13). Notice, too, that Eq. (2.3-13) is a linear feedback law, as promised.

Problem 2.3-2 allows removal of Temporary Assumption 2.3-1, and, accordingly, we may summarize the results as follows.

> **Solution of the regulator problem.** The optimal performance index for the regulator problem with initial time t and initial state $x(t)$ is $x'(t)P(t)x(t)$, where $P(t)$ is given by the solution of the Riccati equation (2.3-11) with initial condition (2.3-12). The matrix $P(t)$ exists for all $t \leq T$. The optimal control for the regulator problem with arbitrary initial time is given by the linear feedback law (2.3-13) for any time t in the interval over which optimization is being carried out.

To illustrate the previous concepts, we consider the following example. The system equation is

$$\dot{x} = \tfrac{1}{2}x + u$$

and the performance index is

$$\int_{t_0}^{T} (2e^{-t}u^2 + \tfrac{1}{2}e^{-t}x^2)\, dt.$$

We require, of course, the optimal control and associated optimal performance index. The Riccati equation associated with this problem is

$$-\dot{P} = P - \tfrac{1}{2}e^{t}P^2 + \tfrac{1}{2}e^{-t} \qquad P(T) = 0.$$

The solution of this equation may be verified to be†

$$P(t) = (1 - e^{t}e^{-T})(e^{t} + e^{2t}e^{-T})^{-1}.$$

†Later, we shall discuss techniques for deriving analytic solutions for Riccati equations.

The optimal control is thus

$$u(t) = -\tfrac{1}{2}(1 - e^t e^{-T})(1 + e^t e^{-T})^{-1}x(t)$$

and the optimal performance index is

$$V^*(x(t_0), t_0) = x'(t_0)(1 - e^{t_0}e^{-T})(e^{t_0} + e^{2t_0}e^{-T})^{-1}x(t_0).$$

Problem 2.3-1. Show that Eqs. (2.3-3) and (2.3-4) imply the existence of a matrix $P(t)$ such that $V^*(x(t), t) = x'(t)P(t)x(t)$. [*Hint:* Without loss of generality, $P(t)$ is symmetric. Use $x = e_i$ in (2.3-3) to define the diagonal entries of $P(t)$, and then specific values of x_1 and x_2 in (2.3-4) to define the off-diagonal entries of $P(t)$. With $P(t)$ so defined, check that the definition is compatible with (2.3-3) and (2.3-4) holding for all x, x_1, and x_2.]

Problem 2.3-2. Consider the regulator problem posed at the start of the section, without Temporary Assumption 2.3-1, and let $P(t)$ be defined by (2.3-11) and (2.3-12). [Continuity of the entries of $F(t)$, etc., is sufficient to guarantee existence of a unique solution in a neighborhood of T.] Define the control law (*not* known to be optimal)

$$u^{**}(t) = -R^{-1}(t)G'(t)P(t)x(t)$$

and show that

$$V(x(t_0), u(\cdot), t_0) = x'(t_0)P(t_0)x(t_0) + \int_{t_0}^{T} (u - u^{**})'R(u - u^{**})\, d\tau.$$

Conclude that if $P(t)$ exists for all $t \in [t_0, T]$, the optimal control is, in fact, u^{**}.

Problem 2.3-3. Find the optimal control for the system (with scalar u and x)

$$\dot{x} = u \qquad x(t_0) \text{ given}$$

and with performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} (u^2 + x^2)\, dt + x^2(T).$$

Problem 2.3-4. Find the optimal control for the system (with scalar u and x)

$$\dot{x} = x + u \qquad x(t_0) \text{ given}$$

and with performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} (u^2 + 2x^2)\, dt.$$

Problem 2.3-5. Another version of the regulator problem imposes a further constraint—namely, that $x(T) = 0$ at the final time T. A class of performance indices implies intrinsic satisfaction of this constraint:

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} (u'Ru + x'Qx)\, dt + \lim_{n\to\infty} nx'(T)x(T).$$

Show that if P satisfies the Riccati equation

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q,$$

and if P^{-1} exists, then P^{-1} satisfies a Riccati equation. Then show that $P^{-1}(T) = 0$ gives a local solution to the Riccati equation for P^{-1}, which, on account of the continuous dependence of solutions of differential equations on the boundary condition, defines the optimal performance index as $x'(t)P(t, T)x(t)$ for t suitably close to T. Then show that this formula is valid for all $t < T$.

REFERENCES

[1] Kalman, R. E., Y. Ho, and K. S. Narendra, "Controllability of Linear Dynamical Systems," *Contributions to Differential Equations*, Vol. 1, John Wiley & Sons, Inc., New York, 1963.

[2] Athans, M., and P. L. Falb, *Optimal Control*, McGraw-Hill Book Company, New York, 1966.

[3] Sage, A. P., *Optimal Systems Control*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968.

[4] Kalman, R. E., "Contributions to the Theory of Optimal Control," *Bol. Soc. Matem. Mex.*, 1960, pp. 102–119.

[5] Kalman, R. E., "The Theory of Optimal Control and the Calculus of Variations," Ch. 16 of *Mathematical Optimization Techniques*, R. Bellman (ed.), University of California Press, 1963.

[6] Kalman, R. E., "When is a Linear Control System Optimal?" *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 86, March 1964, pp. 1–10.

[7] Letov, A. M., "Analytical Controller Design, I," *Automation and Remote Control*, Vol. 21, No. 4, April 1960, pp. 303–306.

[8] Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko, *The Mathematical Theory of Optimal Processes*, K. N. Trirogoff (transl.), L. W. Neustadt (ed.), Interscience Publishers, New York, 1962.

[9] Dreyfus, S. E., *Dynamic Programming and the Calculus of Variations*, Academic Press Inc., New York, 1965.

[10] Bellman, R. E., and S. E. Dreyfus, *Applied Dynamic Programming*, Princeton University Press, Princeton, N. J., 1962.

[11] Bellman, R., and R. Bucy, "Asymptotic Control Theory," *SIAM J. Control*, Vol. 2, No. 1, 1964, pp. 11–18.

[12] Faurre, P. A., "Sur les points conjugués en commande optimale," *Compt. Rend. Acad. Sci. Paris, Ser. A*, Vol. 266, June 24, 1968, pp. 1294–1296.

CHAPTER 3

THE STANDARD REGULATOR PROBLEM—II

3.1 THE INFINITE-TIME REGULATOR PROBLEM

In this section, we lift the restriction imposed in the last chapter that the final time (the right-hand endpoint of the optimization interval), T, be finite. We thus have the following problem.

Infinite-time regulator problem. Consider the system

$$\dot{x} = F(t)x(t) + G(t)u(t) \qquad x(t_0) \text{ given} \tag{3.1-1}$$

with the entries of $F(t)$, $G(t)$ assumed continuous. Let the matrices $Q(t)$ and $R(t)$ have continuous entries, be symmetric, and be nonnegative and positive definite, respectively. Define the performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u'(t)R(t)u(t) + x'(t)Q(t)x(t))\, dt \tag{3.1-2}$$

and the minimization problem as the task of finding an optimal control $u^*(t)$, $t \geq t_0$, minimizing V and the associated optimum performance index $V^*(x(t_0), t_0)$.

It is not always possible to solve this problem as it is stated. To give some idea of the difficulty, consider the system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}x_1\\ x_2\end{bmatrix}+\begin{bmatrix}0\\ 1\end{bmatrix}u \qquad \begin{bmatrix}x_1(t_0)\\ x_2(t_0)\end{bmatrix}=\begin{bmatrix}1\\ 0\end{bmatrix}$$

with performance index defined by $R = [1]$ and

$$Q=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

It is readily established that

$$V(x(t_0), u(\cdot), t)=\int_{t_0}^{\infty}(u^2+e^{2t})\,dt.$$

In a sense, V is minimized by taking $u = 0$; but the resulting value of V is still infinite.

With the finite-time problem, the optimal V is always finite; this may not be so in the infinite-time case. For the example given, it is clear that V becomes infinite for the following three reasons:

1. The state $x_1(t_0)$ is uncontrollable.
2. The uncontrollable part of the system trajectory is unstable ($x_1(t) = e^t$).
3. The unstable part of the system trajectory is reflected in the system performance index (e^{2t} is integrated).

The difficulty would not have arisen if one of these reasons did not apply.

It is intuitively clear that any regulator problem where one had a situation corresponding to that described by (1), (2), and (3) cannot have a finite optimal performance index. Therefore, to ensure that our problems are solvable, we shall make the following assumption.

ASSUMPTION 3.1-1. System (3.1-1) is completely controllable for every time t. That is, given an arbitrary state $x(t)$ at time t, there exists a control depending on $x(t)$ and t, and a time t_2 depending on t, such that application of this control over the interval $[t, t_2]$ takes the state $x(t)$ to the zero state at time t_2.

In the problems at the end of the section, we shall indicate a variant upon this assumption that will also ensure solvability of the infinite-time regulator problem.

We now state the solution to the infinite-time regulator problem, under Assumption 3.1-1.

Solution to the infinite-time regulator problem. Let $P(t, T)$ be the solution of the equation

$$-\dot{P}=PF+F'P-PGR^{-1}G'P+Q \tag{3.1-3}$$

with boundary condition $P(T, T) = 0$. Then $\lim_{T\to\infty} P(t, T) = \bar{P}(t)$ exists for all t and is a solution of (3.1-3). Moreover, $x'(t)\bar{P}(t)x(t)$ is the optimal

performance index, when the initial time is t and the initial state is $x(t)$. The optimal control at time t when the initial time is arbitrary is uniquely† given by

$$u^*(t) = -R^{-1}(t)G'(t)\bar{P}(t)x(t) \tag{3.1-4}$$

(assuming t lies in the optimization interval).

Evidently, we need to establish four separate results: (1) the existence of $\bar{P}(t)$; (2) the fact that it is a solution of (3.1-3); (3) the formula for the optimal performance index; and (4) the formula for the optimal control. The proof of (3) and (4) will be combined.

Existence of $\bar{P}(t)$. Since (3.1-1) is completely controllable at time t, there exists for every $x(t)$ a control $\tilde{u}(\cdot)$ and a time t_2 such that $\tilde{u}(\cdot)$ transfers $x(t)$ to the zero state at time t_2. Although $\tilde{u}(\cdot)$ is initially defined only on $[t, t_2]$, we extend the definition to $[t, \infty)$ by taking $\tilde{u}(\cdot)$ to be zero after t_2. This ensures that the system will remain in the zero state after time t_2. The notation $V(x(t), u(\cdot), t, T)$ will be used to denote the performance index resulting from initial state $x(t)$ at time t, a control $u(\cdot)$, and a final time T, which is finite rather than infinite as in (3.1-2). Then $P(t, T)$ exists for all T and $t \le T$. Moreover,

$$\begin{aligned} x'(t)P(t, T)x(t) &= V^*(x(t), t, T) \\ &\le V(x(t), \tilde{u}_{[t,T]}, t, T) \\ &\le V(x(t), \tilde{u}_{[t,\infty)}, t, \infty) \\ &= V(x(t), \tilde{u}_{[t,t_2]}, t, t_2) \\ &< \infty. \end{aligned}$$

Since $x(t)$ is arbitrary, it follows that the entries of $P(t, T)$ are bounded independently of T. [Note that as T approaches infinity, if any entry of $P(t, T)$ became unbounded, there would have to be a diagonal entry of $P(t, T)$ which became unbounded, tending to infinity; consequently, for a suitable $x(t)$, $x'(t)P(t, T)x(t)$ would be unbounded.]

Reference to (3.1-2), with T replacing $+\infty$ on the integral, and use of the nonnegative definite and positive definite character of Q and R, respectively, show that

$$x'(t)P(t, T_0)x(t) \le x'(t)P(t, T_1)x(t)$$

for any $T_1 \ge T_0$. The bound on $P(t, T)$ together with this monotonicity relation guarantees existence of the limit $\bar{P}(t)$. In more detail, the existence

†At least, $u^*(t)$ is uniquely defined up to a set of measure zero, unless we insist on some property such as continuity. Henceforth, we shall omit reference to this qualification. (Those unfamiliar with measure theory may neglect this point.)

of the limit $\bar{p}_{ii}(t)$ will follow by taking $x(t) = e_i$, for each i, and applying the well-known result of analysis that a bounded monotonically increasing function possesses a limit. (Recall that e_i is defined as a vector with zeros for all entries except the ith, where the entry is 1.) The existence of $\bar{p}_{ij}(t)$ will follow by observing that $2p_{ij}(t, T) = (e_i + e_j)'P(t, T)(e_i + e_j) - p_{ii}(t, T) - p_{jj}(t, T)$, and that each term on the right possesses a limit as T approaches infinity.

$\bar{P}(t)$ satisfies the Riccati equation (3.1-3). Denote the solution of (3.1-3) satisfying $P(T) = A$ by $P(t, T; A)$. [Then $P(t, T)$, defined earlier, is $P(t, T; 0)$.] Now a moment's reflection shows that

$$P(t, T; 0) = P(t, T_1; P(T_1, T; 0))$$

for $t \leq T_1 \leq T$, and thus

$$\bar{P}(t) = \lim_{T\to\infty} P(t, T; 0) = \lim_{T\to\infty} P(t, T_1; P(T_1, T; 0).$$

For fixed time T_1, the solution $P(t, T_1; A)$ of (3.1-3) depends continuously on A; therefore,

$$\begin{aligned}\bar{P}(t) &= P(t, T_1; \lim_{T\to\infty} P(T_1, T; 0)) \\ &= P(t, T_1; \bar{P}(T_1)),\end{aligned}$$

which proves that $\bar{P}(t)$ is a solution of (3.1-3) defined for all t.

Optimal performance index and control formulas. We show first that if the control defined by (3.1-4) is applied (where there is no assumption that this control is optimal), then

$$V(x(t), u^*(\cdot), t, \infty) = \lim_{T\to\infty} V(x(t), u^*(\cdot), t, T) = x'(t)\bar{P}(t)x(t). \tag{3.1-5}$$

Direct substitution of (3.1-4) into the performance index (3.1-2), with the initial time replaced by t and the final time by T, leads to

$$\begin{aligned}V(x(t), u^*(\cdot), t, T) &= x'(t)\bar{P}(t)x(t) - x'(T)\bar{P}(T)x(T) \\ &\leq x'(t)\bar{P}(t)x(t),\end{aligned}$$

and, therefore,

$$\lim_{T\to\infty} V(x(t), u^*(\cdot), t, T) \leq x'(t)\bar{P}(t)x(t).$$

Also,

$$\begin{aligned}V(x(t), u^*(\cdot), t, T) &\geq V^*(x(t), t, T) \\ &\geq x'(t)P(t, T)x(t)\end{aligned}$$

and, therefore,

$$\lim_{T\to\infty} V(x(t), u^*(\cdot), t, T) \geq x'(t)\bar{P}(t)x(t).$$

The two inequalities for $\lim_{T\to\infty} V(x(t), u^*(\cdot), t, T)$ then imply (3.1-5). Since

$u^*(\cdot)$ has not yet been shown to be optimal, it follows that

$$V^*(x(t), t, \infty) \leq V(x(t), u^*(\cdot), t, \infty). \tag{3.1-6}$$

We now show that the inequality sign is impossible. Suppose that strict inequality holds. Then there is a control u_1, different from u^*, such that

$$\lim_{T\to\infty} V(x(t), u_1, t, T) = V^*(x(t), t, \infty).$$

Since, also, from (3.1-5),

$$\lim_{T\to\infty} V^*(x(t), t, T) = V(x(t), u^*, t, \infty),$$

it follows that strict inequality in (3.1-6) implies

$$\lim_{T\to\infty} V(x(t), u_1, t, T) < \lim_{T\to\infty} V^*(x(t), t, T).$$

This, in turn, requires for suitably large T

$$V(x(t), u_1, t, T) < V^*(x(t), t, T).$$

This is plainly impossible by the definition of the optimal performance index as the minimum over all possible indices. Consequently, we have established that $x'(t)\bar{P}(t)x(t)$ is the optimal performance index for the infinite-time problem, and that $-R^{-1}(t)G'(t)\bar{P}(t)x(t)$ is the unique optimal control because it achieves this performance index, thereby completing the formal solution to the infinite-time regulator problem.

It is of interest, for practical reasons, to determine whether time-invariant plants (3.1-1) will give rise to time-invariant linear control laws of the form

$$u(t) = K'x(t). \tag{3.1-7}$$

For finite-time optimization problems of the type considered in the last chapter, no choice of T, $R(\cdot)$, and $Q(\cdot)$ will yield a time-invariant control law when F and G are constant, unless the matrix A takes on certain special values. [Problem 3.1-2 asks for this fact to be established.] For the infinite-time problem, the case is a little different. Let us state an infinite-time problem that will be shown to yield a constant control law. In the next chapter, we shall consider a variation on this problem statement which also yields a constant (and linear) control law.

Time-invariant regulator problem. Consider the system

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{3.1-8}$$

where F and G are constant. Let the constant matrices Q and R be nonnegative and positive definite, respectively. Define the performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u'Ru + x'Qx)\, dt \tag{3.1-9}$$

and the minimization problem as the task of finding an optimal control

$u^*(\cdot)$ minimizing V, together with the associated optimum performance index.

To ensure solvability of the problem, it is necessary, as before, to make some additional restrictions. We shall require the following specialization of Assumption 3.1-1.

ASSUMPTION 3.1-2. The system (3.1-8) is completely controllable.

Since the system (3.1-8) is time invariant, there is no distinction between complete controllability at all times and complete controllability at some particular time.

Solution to the time-invariant regulator problem. Let $P(t, T)$ be the solution of the equation

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \tag{3.1-3}$$

with initial condition $P(T, T) = 0$. Then $\lim_{T\to\infty} P(t, T) = \bar{P}$ exists and is constant; also, $\bar{P} = \lim_{t\to-\infty} P(t, T)$. Furthermore, $\bar{P}$ satisfies (3.1-3); that is,

$$\bar{P}F + F'\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q = 0 \tag{3.1-10}$$

and $x'(t)\bar{P}x(t)$ is the optimal performance index when the initial time is t and the initial state is $x(t)$. The optimal control at time t when the initial time is arbitrary is uniquely given by the constant control law

$$u^*(t) = -R^{-1}G'\bar{P}x(t). \tag{3.1-11}$$

Given the time-varying results, it is quite straightforward to establish the various claims just made.

First, $\lim_{T\to\infty} P(t, T)$ certainly exists. Now the plant is time invariant, and the function under the integral sign of the performance index is not specifically time dependent. This means that the choice of the initial time is arbitrary—i.e., all initial times must give the same performance index, which is to say that $\bar{P}$ is constant. Likewise, because the initial time is arbitrary,

$$\bar{P} = \lim_{T\to\infty} P(t, T) = \lim_{T\to\infty} P(0, T - t) = \lim_{t\to-\infty} P(0, T - t) = \lim_{t\to-\infty} P(t, T).$$

For the (time-varying) infinite-time problem, it was established that $\bar{P}(t)$ satisfies the Riccati equation (3.1-3). Consequently, the constant $\bar{P}$ here satisfies (3.1-3); since $\dot{\bar{P}}$ is now zero, (3.1-10) follows. (It is sometimes said that $\bar{P}$ is a *steady state solution* of the Riccati equation.)

The remainder of the claims are immediate consequences of the more general results applying for the time-varying problem.

To illustrate the preceding concepts, we can consider the following

simple example. The prescribed system is

$$\dot{x} = x + u$$

and the performance index is

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u^2 + x^2)\, dt.$$

To find the optimal control, we solve

$$-\dot{P} = 2P - P^2 + 1 \qquad P(T, T) = 0.$$

It is readily found that

$$P(t, T) = \frac{\exp[-2\sqrt{2}(t - T)]}{(\sqrt{2} + 1) + (\sqrt{2} - 1)\exp[-2\sqrt{2}(t - T)]}$$

and

$$\lim_{t \to -\infty} P(t, T) = \bar{P} = \sqrt{2} + 1.$$

(Observe, also, that $0 = 2\bar{P} - \bar{P}^2 + 1$, as required. Indeed, this equation could have been used to determine $\bar{P}$, assuming we knew which of the two solutions of the equation to choose.) The optimal control is

$$u^* = -(\sqrt{2} + 1)x$$

(and thus the closed-loop system becomes $\dot{x} = -\sqrt{2}x$).

A second example is provided by a voltage regulator problem, discussed in [1] and [2]. The open-loop plant consists of a cascade of single-order blocks, of transfer functions

$$\frac{3}{0.1s + 1}, \frac{3}{0.04s + 1}, \frac{6}{0.07s + 1}, \frac{3.2}{2s + 1}, \frac{2.5}{5s + 1}.$$

In other words, the Laplace transform $Y(s)$ of the deviation from the correct output is related to the Laplace transform $U(s)$ of the deviation from the reference input by

$$Y(s) = \frac{3}{0.1s + 1} \cdot \frac{3}{0.04s + 1} \cdot \frac{6}{0.07s + 1} \cdot \frac{3.2}{2s + 1} \cdot \frac{2.5}{5s + 1} U(s).$$

State-space equations relating $u(\cdot)$ to $y(\cdot)$ are readily found to be

$$\dot{x} = \begin{bmatrix} -0.2 & 0.5 & 0 & 0 & 0 \\ 0 & -0.5 & 1.6 & 0 & 0 \\ 0 & 0 & -\frac{1}{7} & \frac{6}{7} & 0 \\ 0 & 0 & 0 & -0.25 & 7.5 \\ 0 & 0 & 0 & 0 & -0.1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0.3 \end{bmatrix} u$$

[Here, $x_1, x_2, \ldots, x_5$ have physical significance; x_5 is the output of a block of transfer function $3(0.1s + 1)^{-1}$ with input u, x_4 the output of a block of transfer function $3(0.4s + 1)^{-1}$ with input x_5, etc. Also, $x_1 = y$.]

As a first attempt at optimization, the performance index

$$\int_{t_0}^{\infty} (x_1^2 + u^2)\, dt$$

is considered. (Of course, $x_1^2 = x'Qx$, where Q is a matrix the only nonzero entry of which is unity in the 1–1 position.)

Solution of the appropriate Riccati equation requires procedures discussed subsequently in this book. The important quantity calculated is not so much the matrix $\bar{P}$ but the optimal gain vector. The optimal control law is, in fact, found to be

$$u = -[0.9243 \quad 0.1711 \quad 0.0161 \quad 0.0392 \quad 0.2644]x.$$

This problem also emphasizes the fact that the regulator theory is applicable to systems where the state is to be controlled around a nominal nonzero value, with a control varying around an associated nominal nonzero value.

Problem 3.1-1. Consider the time-invariant problem, and suppose F and G have the form

$$F = \begin{bmatrix} F_{11} & F_{12} \\ 0 & F_{22} \end{bmatrix} \qquad G = \begin{bmatrix} G_1 \\ 0 \end{bmatrix}$$

so that the pair $[F, G]$ is not completely controllable. [One way to see this is to examine the rank of $[G \quad FG \quad \ldots \quad]$; see Appendix B.] Show that if the eigenvalues of F_{22} have negative real part, it is still possible to find a solution to the infinite-time regulator problem. Suppose that the solution $P(t)$ of the Riccati equation associated with the regulator problem is partitioned as

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P'_{12} & P_{22} \end{bmatrix}.$$

Show that P_{11} satisfies a Riccati equation, that P_{22} satisfies a linear differential equation—assuming P_{11} is known—and that P_{12} satisfies a linear differential equation—assuming P_{22} is known. (This problem is discussed in [3].)

Problem 3.1-2. Consider the finite-time regulator problem posed in the previous section, where the linear system is time invariant. Show that if $A = 0$, there is no way, even permitting time-varying $R(t)$ and $Q(t)$, of achieving a constant control law. Show that if R and Q are constant, then there is at least one particular A that will achieve a constant control law, and that this A is independent of the final time T. [*Hint:* For the second part of the problem, make use of Eq. (3.1-10).]

Problem 3.1-3. Find the optimum performance index and optimal control for the system $\dot{x} = ax + u$ with performance index $\int_{t_0}^{\infty} (u^2 + bx^2)\, dt$, where a and b are constant and b is positive. Write down the equation of the closed-loop system, and show that the closed-loop system is always asymptotically stable.

3.2 STABILITY OF THE TIME-INVARIANT REGULATOR

In this section, we shall be concerned with the stability of the closed-loop system formed when the control law resulting from an infinite-time performance index is implemented. Throughout, we shall assume a constant system that is completely controllable:

$$\dot{x} = Fx + Gu. \tag{3.2-1}$$

We shall also assume a performance index with constant nonnegative definite Q and constant positive definite R:

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u'Ru + x'Qx)\, dt. \tag{3.2-2}$$

We recall that the optimal performance index is $x'(t_0)\bar{P}x(t_0)$, where $\bar{P}$ is the limiting solution of a Riccati equation; also, $\bar{P}$ satisfies the algebraic equation

$$\bar{P}F + F'\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q = 0. \tag{3.2-3}$$

The optimal control is given by

$$u^* = -R^{-1}G'\bar{P}x, \tag{3.2-4}$$

and, accordingly, the closed-loop system becomes

$$\dot{x} = (F - GR^{-1}G'\bar{P})x. \tag{3.2-5}$$

We ask the question: When is (3.2-5) an asymptotically stable system? Certainly, (3.2-5) is not always stable. Consider the example $\dot{x} = x + u$, with $V = \int_{t_0}^{\infty} u^2\, dt$. Immediately, the optimal control is $u \equiv 0$, and the closed-loop system is $\dot{x} = x$, which is plainly unstable. In this instance, there are two factors contributing to the difficulty.

1. The original open-loop system is unstable.
2. The unstable trajectories do not contribute in any way to the performance index—in a sense, the unstable states are not *observed* by the performance index.

Intuitively, one can see that if (1) and (2) were true in an arbitrary optimization problem, there would be grounds for supposing that the closed-loop system would be unstable.

Accordingly, to ensure asymptotic stability of the closed-loop system, it is necessary to prevent the occurrence of (1) and (2). This motivates the introduction of the following assumption, which will be shown to guarantee stability of the closed-loop system (see also Problem 3.2-1 for an alternative assumption).

ASSUMPTION 3.2-1. The pair $[F, D]$ is *completely observable*, where D is any matrix such that $DD' = Q$. (Equivalently, the equation $D'e^{Ft}x_0 = 0$ for all t implies $x_0 = 0$, see Appendix B.)

We can note at once that the question of whether Assumption (3.2-1) holds is determined by Q alone, and not by the particular factorization DD'. To see this, suppose D_1 and D_2 are such that $D_1D_1' = D_2D_2' = Q$, with $[F, D_1]$ completely observable and $[F, D_2]$ not completely observable. Then there exists a nonzero x_0 such that $D_2'e^{Ft}x_0 = 0$ for all t, and we have

$$\begin{aligned} D_2'e^{Ft}x_0 = 0 \quad \text{for all } t &\Longrightarrow x_0'e^{F't}D_2D_2'e^{Ft}x_0 = 0 \quad \text{for all } t \\ &\Longrightarrow x_0'e^{F't}D_1D_1'e^{Ft}x_0 = 0 \quad \text{for all } t \\ &\Longrightarrow D_1'e^{Ft}x_0 = 0 \quad \text{for all } t \end{aligned}$$

which contradicts the complete observability of $[F, D_1]$. Hence, either $[F, D_1]$ and $[F, D_2]$ are completely observable simultaneously, or they are not completely observable simultaneously.

Assumption 3.2-1 essentially ensures that all trajectories will show up in the $x'Qx$ part of the integrand of the performance index. Since the performance index is known a priori to have a finite value, it is plausible that any potentially unstable trajectories will be stabilized by the application of the feedback control.

The actual proof of asymptotic stability requires us first to note the following result.

LEMMA. Consider the time-invariant regulator problem as defined by Eqs. (3.2-1) through (3.2-5) and the associated remarks. Then Assumption 3.2-1 is necessary and sufficient for $\bar{P}$ to be symmetric positive definite.

Proof. From the form of the performance index (3.2-2), it is clear that $x'(t_0)\bar{P}x(t_0)$ must be nonnegative for all $x(t_0)$. Suppose for some nonzero x_0 we have $x_0'\bar{P}x_0 = 0$, with Assumption 3.2-1 holding. Now the only way the integral in (3.2-2) will turn out to be zero is if the nonnegative integrand is always zero. This requires the optimal control to be zero for all t; consequently, the system state at time t when the optimal control is applied is simply $\exp[F(t - t_0)]x_0$. Furthermore,

$$0 = x_0'\bar{P}x_0 = \int_{t_0}^{\infty} x'Qx\,dt = \int_{t_0}^{\infty} x_0' \exp[F'(t - t_0)]DD' \exp[F(t - t_0)]x_0\,dt.$$

Now, Assumption 3.2-1 leads to a contradiction, because the preceding equation implies $D'\exp[F(t - t_0)]x_0 = 0$ for all t, and we have assumed x_0 to be nonzero.

To establish the converse, suppose there exists a nonzero x_0 such that $D'\exp[F(t - t_0)]x_0 = 0$ for all t. Apply the control $u(t) = 0$ for all t to the open-loop plant. It then follows that the associated performance index is zero. Since the optimal performance index is bounded below by zero, it follows that it, too, is zero—i.e., $x_0'\bar{P}x_0 = 0$, which implies finally that $\bar{P}$ is singular.

Before applying the lemma, we recall two results in the theory of Lyapunov stability. The first is well-known, whereas the second is perhaps not so familiar but may be found in [4] and [5].

THEOREM A. Consider the *time-invariant* system $\dot{x} = f(x)$, with $f(0) = 0$ and $f(\cdot)$ continuous. Suppose there is a scalar function $V(x)$ that is positive definite, approaches infinity as $\|x\|$ approaches infinity, and is differentiable. Suppose also that the derivative of V along system trajectories—namely, $\dot{V} = [\partial V/\partial x]'f(x)$—is negative definite. Then the system is globally asymptotically stable.

For the second theorem, we relax the negative definiteness assumption on $\dot{V}$. A well-known result yields that if $\dot{V}$ is nonpositive definite, stability rather than asymptotic stability prevails. But by sharpening the constraint on $\dot{V}$ slightly, we have the following.

THEOREM B. Consider the *time-invariant* system $\dot{x} = f(x)$, with $f(0) = 0$ and $f(\cdot)$ continuous. Suppose there is a scalar function $V(x)$ that is positive definite, approaches infinity as $\|x\|$ approaches infinity, and is differentiable. Suppose also that the derivative of V along system trajectories—namely, $\dot{V} = [\partial V/\partial x]'f(x)$—is nonpositive definite, and not identically zero on $[t_1, \infty]$ for any t_1, save for a trajectory starting in the zero state. Then the system is globally asymptotically stable.

The proof of asymptotic stability of the closed-loop system (3.2-5) is now straightforward. From (3.2-3), it follows that

$$\bar{P}(F - GR^{-1}G'\bar{P}) + (F' - \bar{P}GR^{-1}G')\bar{P} = -Q - \bar{P}GR^{-1}G'\bar{P}. \qquad (3.2\text{-}6)$$

Take $V(x) = x'\bar{P}x$ as a prospective Lyapunov function for the closed-loop system (3.2-5). Then the lemma ensures that $V(x)$ is positive definite, whereas (3.2-6) implies that

$$\dot{V}(x) = -x'Qx - x'\bar{P}GR^{-1}G'\bar{P}x. \qquad (3.2\text{-}7)$$

Now, if Q is positive definite, as distinct from merely nonnegative definite, then $\dot{V}$ is negative definite, and thus we have asymptotic stability from Theorem *A*.

Suppose now Q is nonnegative definite [of course, Assumption (3.2-1) remains in force]. Certainly $\dot{V}$ is nonpositive, by inspection. To conclude asymptotic stability using Theorem B, we must show that $\dot{V}$ is not identically zero along system trajectories starting from nonzero initial states.

Suppose $\dot{V}$ is identically zero along a trajectory starting from a nonzero initial state x_0. Then $x'\bar{P}GR^{-1}G'\bar{P}x$ is identically zero; furthermore, $-R^{-1}G'\bar{P}x(t)$, which is the (optimal) control for the open-loop system, is also identically zero. Therefore, the trajectories of the closed-loop system are the same as those of the open-loop system and are given from $x(t) =$

$\exp[F(t - t_0)]x_0$. Because the assumption $\dot{V}$ is identically zero also implies that $x'Qx$ is identically zero, then $x_0' \exp[F'(t - t_0)]DD' \exp[F(t - t_0)]x_0$ must be identically zero. This contradicts Assumption 3.2-1 or the fact that $x_0 \neq 0$. Consequently, it is impossible to have $\dot{V}$ identically zero along a trajectory, other than that starting at the zero state. The asymptotic stability of (3.2-5) is thus established.

The practical implications of the stability result should be clear. Normally no one wants an unstable system; here, we have a procedure for guaranteeing that an optimal system is bound to be stable, irrespective of the stability of the open-loop plant. The result is in pleasant contrast to some of the results and techniques of classical control, where frequently the main aim is to achieve stability and questions of optimality occupy an essentially secondary role in the design procedure. The contrast will actually be heightened when we exhibit some of the additional virtues of the optimal regulator solution in later chapters.

The interested reader may wonder to what extent the stability result applies in time-varying, infinite-time problems. In general, further assumptions than merely Assumption 3.2-1 are needed, as discussed in Sec. 14.2; reference [6] should also be consulted. On the other hand, reference [7] gives in some detail the theory of the time-invariant problem.

To illustrate these ideas, we recall briefly the two examples of the previous section. In the first case, we had

$$\dot{x} = x + u \qquad V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u^2 + x^2)\, dt.$$

Obviously, the observability condition is satisfied. Therefore, it is no surprise that the closed-loop system, computed to be $\dot{x} = -\sqrt{2}\,x$, is asymptotically stable.

For the voltage regulator example, we recall that the F matrix is

$$\begin{bmatrix} -0.2 & 0.5 & 0 & 0 & 0 \\ 0 & -0.5 & 1.6 & 0 & 0 \\ 0 & 0 & -\frac{1}{7} & \frac{6}{7} & 0 \\ 0 & 0 & 0 & -0.25 & 7.5 \\ 0 & 0 & 0 & 0 & -0.1 \end{bmatrix}$$

and the weighting matrix Q is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A matrix D such that $DD' = Q$ is given by $D' = [1 \quad 0 \quad 0 \quad 0 \quad 0]$. To check the observability condition, the easiest way, see Appendix B, is to examine the rank of $[D \quad F'D \quad \cdots \quad (F')^4 D]$. This matrix is readily checked to be triangular with nonzero elements all along the diagonal. Hence, it has rank 5. Consequently, the closed-loop system is known to be asymptotically stable.

This could also be checked, with considerable effort, by evaluating the characteristic polynomial associated with the closed-loop system, and then checking via the Routh or other procedure that the roots of the polynomial possess negative real parts. There is little point in such a calculation here, however.

Problem 3.2-1. With quantities as defined in Eqs. (3.2-1) through (3.2-5), suppose also that

$$F = \begin{bmatrix} F_{11} & F_{12} \\ 0 & F_{22} \end{bmatrix} \quad \text{and} \quad D' = [0 \quad D'_{22}] \quad \text{with} \quad DD' = Q.$$

(Then $[F, D]$ is not completely observable.) Show that the closed-loop system is asymptotically stable if all eigenvalues of F_{11} have negative real parts.

Problem 3.2-2. The scalar version of (3.2-3) is a simple quadratic equation, which, in general, has more than one solution. Likewise, when (3.2-3) is a true matrix equation, there is generally more than one solution. Show that there is only one positive definite solution under Assumption 3.2-1. [This identifies the limiting solution of the Riccati equation uniquely as *the* positive definite solution of (3.2-3).] {*Hint:* Suppose there are two solutions, P_1 and P_2, both positive definite. Define $F_i = F - GR^{-1}G'P_i$ and show that both F_i have eigenvalues with negative real parts. Prove that $(P_1 - P_2)F_1 + F'_2(P_1 - P_2) = 0$ and use the result of [8] that the matrix equation $AX + XB = 0$ has the unique solution $X = 0$, provided $\lambda_i[A] + \lambda_j[B] \neq 0$ for any i and j; see also Appendix A for this result.}

Problem 3.2-3. Can you state some explicit conditions under which the closed-loop system will fail to be asymptotically stable?

Problem 3.2-4. Consider again Problem 3.1-3 of this chapter. Is there any relation between the exponential rate at which $P(t, T)$ approaches $\bar{P}$ as t approaches minus infinity and the exponential rate of decay of a nonzero initial state in the closed-loop system?

3.3 SUMMARY AND DISCUSSION OF THE REGULATOR PROBLEM RESULTS

In this section, we first summarize all the important regulator problem results hitherto established.

Regulator problem and solution. Consider the system

$$\dot{x} = F(t)x + G(t)u \qquad x(t_0) \text{ given} \tag{3.3-1}$$

with the entries of F and G assumed continuous. Let the matrices $Q(t)$ and $R(t)$ have continuous entries, be symmetric, and be nonnegative and positive definite, respectively. Let A be a nonnegative definite matrix. Define the performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} (u'(t)R(t)u(t) + x'(t)Q(t)x(t))\,dt + x'(T)Ax(T) \tag{3.3-2}$$

where T is finite. Then the minimum value of the performance index is

$$V^*(x(t_0), t_0) = x'(t_0)P(t_0, T)x(t_0) \tag{3.3-3}$$

where $P(t, T)$ is the solution of the Riccati equation

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \tag{3.3-4}$$

with boundary condition $P(T, T) = A$. The matrix $P(t, T)$ exists for all $t \leq T$. The associated optimal control is given by the linear feedback law

$$u^*(t) = -R^{-1}(t)G'(t)P(t, T)x(t). \tag{3.3-5}$$

Infinite-time regulator problem and solution. Suppose the preceding hypotheses all hold, save that $A = 0$. Suppose also that the system (3.3-1) is completely controllable for all time. Then

$$\bar{P}(t) = \lim_{T\to\infty} P(t, T) \tag{3.3-6}$$

exists, and the minimum value of the performance index (3.3-2) with T replaced by infinity is $x'(t_0)\bar{P}(t_0)x(t_0)$. The matrix $\bar{P}$ satisfies the Riccati equation (3.3-4), and the optimal control law is

$$u^*(t) = -R^{-1}(t)G'(t)\bar{P}(t)x(t). \tag{3.3-7}$$

Time-invariant regulator problem and solution. Suppose the hypotheses required for the infinite-time regulator problem and solution are present, and suppose also that the matrices F, G, Q, and R are constant. Then

$$\bar{P} = \lim_{T\to\infty} P(t, T) = \lim_{t\to-\infty} P(t, T) \tag{3.3-8}$$

is constant, and satisfies the algebraic equation

$$\bar{P}F + F'\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q = 0. \tag{3.3-9}$$

The optimal control law is a time-invariant law

$$u^*(t) = -R^{-1}G'\bar{P}x(t). \tag{3.3-10}$$

Asymptotically stable time-invariant regulator problem and solution. Suppose the hypotheses required for the infinite-time regulator problem

and solution are present, and suppose also that the pair $[F, D]$ is completely observable where D is any matrix such that $DD' = Q$. Then $\bar{P}$ is positive definite [and, by Problem (3.2-2), is the only solution of (3.3-9) with this property.] Moreover, the optimal closed-loop system

$$\dot{x} = [F - GR^{-1}G'\bar{P}]x \tag{3.3-11}$$

is asymptotically stable, and $x'\bar{P}x$ is a Lyapunov function.

Now that the solution to the regulator problem has been determined, we might ask if this solution is of interest for application to situations other than those in which everything is set up as in the formal statements just summarized.

Frequently, the system states are not available for use in a controller input. In Chapters 8 and 9, we show that if the system is specified in the form

$$\dot{x} = Fx + Gu \qquad y = H'x, \tag{3.3-12}$$

where only the input u and output y are available for controller inputs, a state estimator can be constructed with inputs u and y and output $\hat{x}$, an estimate of the state x. Implementing the control law $u = -R^{-1}G'P\hat{x}$ is then a satisfactory alternative to implementing the control law $u = -R^{-1}G'Px$.

So far in our discussions, we have assumed that the performance index matrices R and Q are specified. Since specification is usually up to the designer, a range of values for these matrices can be considered and by "trial and error" the most suitable values selected. This approach is often inefficient and leaves something to be desired. Of course, if we are only interested in stability, then we have the result that if the open-loop system $\dot{x} = Fx$ is unstable, the closed-loop system $\dot{x} = (F - GR^{-1}G'\bar{P})x$ is always stable, irrespective of the choice of Q and R within their prescribed limits. In other words, we have a general method for stabilizing multiple input linear systems (assuming that state estimators may be constructed).

We are now led to ask whether linear regulators, optimal in the sense previously discussed, have desirable properties for the engineer other than simply "stability." Or we might ask whether there are methods for selecting the index parameters Q and R so that desirable properties, such as good transient response and good sensitivity characteristics, are achieved.

With these questions in mind, we now move on to the next section and chapter, which discuss various extensions of the regulator problem, and then to the following few chapters, which deal with further properties of regulator systems.

Problem 3.3-1. State the physical significance of choosing $Q = HH'$ and $R = I$ (the unit matrix) in setting up a performance index for the system (3.3-12).

3.4 EXTENSIONS OF THE REGULATOR PROBLEM

There are many avenues along which the regulator problem can be extended; a number of them will be explored in later chapters. We shall content ourselves here by stating a few examples and indicating more completely one simple extension. The following are four examples of extensions to be more fully explored later.

1. Time-invariant systems, with performance indices of the form $\int_{t_0}^{\infty} e^{2\alpha t}(u'Ru + x'Qx)\,dt$, where α, Q, and R are all constant. These will be shown to lead to a constant control law, and to a closed-loop system with improved stability properties.
2. Performance indices of the form $\int_{t_0}^{T} (\dot{u}'R\dot{u} + x'Qx)\,dt$. These lead to feedback laws that are no longer memoryless but dynamic. Regulators designed using this approach can accommodate slowly varying input disturbances.
3. Performance indices of the form $\int_{t_0}^{T} [u'Ru + (x - \tilde{x})'Q(x - \tilde{x})]\,dt$, where $\tilde{x}$ is a prescribed vector function. This is the servomechanism or tracking problem, where one desires the system states to follow a prescribed trajectory $\tilde{x}$.
4. Performance indices of the form $\int_{t_0}^{T} x'Qx\,dt$. For this case, a constraint on the magnitude of u is also included. The resultant optimal system operates in two modes.

The approach used in developing the preceding extensions is to reduce by a suitable transformation the given problem to the standard regulator problem. To illustrate this, we now consider a straightforward example of the technique. One application of this extension will be made when we consider (4).

We consider the determination of an optimal control and associated optimal performance index for the system

$$\dot{x} = F(t)x + G(t)u \qquad x(t_0) \text{ given} \tag{3.4-1}$$

when the performance index is

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} [u'(t)R(t)u(t) + 2x'(t)S(t)u(t) + x'(t)Q(t)x(t)]\,dt \tag{3.4-2}$$

with R positive definite and the following constraint holding:

$$Q - SR^{-1}S' \geq 0 \tag{3.4-3}$$

(shorthand for $Q - SR^{-1}S'$ is nonnegative definite). If desired, T can be infinite, and F, G, Q, R, and S constant.

To reduce this problem to one covered by the previous theory, we note

the following identity, obtained by completing the square:

$$u'Ru + 2x'Su + x'Qx = (u + R^{-1}S'x)'R(u + R^{-1}S'x) + x'(Q - SR^{-1}S')x.$$

Making the definition

$$u_1 = u + R^{-1}S'x, \tag{3.4-4}$$

the original system (3.4-1) becomes equivalent to

$$\dot{x} = (F - GR^{-1}S')x + Gu_1 \tag{3.4-5}$$

and the original performance index is equivalent to

$$V(x(t_0), u_1(\cdot), t_0) = \int_{t_0}^{T} [u_1'Ru_1 + x'(Q - SR^{-1}S')x]\,dt. \tag{3.4-6}$$

If u and u_1 are related according to (3.4-4), the trajectories of the two systems (3.4-1) and (3.4-5) are the same [provided they start from the same initial state $x(t_0)$]. Furthermore, the values taken by the two performance indices—viz., (3.4-2), which is associated with the system (3.4-1), and (3.4-6), which is associated with the system (3.4-5)—are also the same. Consequently, the following statements hold.

1. The optimal controls u^* and u_1^* for the two optimization problems are related by $u_1^* = u^* + R^{-1}S'x$.
2. The optimal performance indices for the two problems are the same.
3. The closed-loop trajectories (when the optimal controls are implemented) are the same.

Now the optimization problem associated with (3.4-5) and (3.4-6) is certainly solvable [and here we are using the nonnegativity constant (3.4-3)]. The optimal u_1^* is given by

$$u_1^*(t) = -R^{-1}(t)G'(t)P(t, T)x(t),$$

where

$$-\dot{P} = P(F - GR^{-1}S') + (F' - SR^{-1}G')P - PGR^{-1}G'P + Q - SR^{-1}S'$$

with $P(T, T) = 0$. The optimal index is $x'(t_0)P(t_0, T)x(t_0)$. The optimal control for the optimization problem associated with (3.4-1) and (3.4-2) is thus

$$u^*(t) = -R^{-1}(t)[G'(t)P(t, T) + S'(t)]x(t)$$

and the optimal performance index is again $x'(t_0)P(t_0, T)x(t_0)$.

To consider the infinite-time problem [i.e., T in (3.4-2) is replaced by infinity], we make the following assumption.

ASSUMPTION 3.4-1. The system (3.4-1) is completely controllable at every time t.

To ensure existence of the limit as T approaches infinity of $P(t, T)$, one evidently requires that the system (3.4-5) be completely controllable at every time t. [Then the optimization problem associated with (3.4-5) and (3.4-6) is solvable and yields a solution of the optimization problem associated with (3.4-1) and (3.4-2).] The complete controllability of (3.4-5) is an immediate consequence of the following lemma.

LEMMA. Suppose the system

$$\dot{x} = F(t)x + G(t)u \tag{3.4-1}$$

is completely controllable at some time t_1. That is, with an arbitrary state $\bar{x}(t_1)$, there is associated a control $\bar{u}(\cdot)$ and time t_2 such that (3.4-1) is transferred by $\bar{u}(\cdot)$ from the state $\bar{x}(t_1)$ at time t_1 to the zero state at time t_2. Then, with $K(t)$ a matrix of continuous functions, the system

$$\dot{x} = [F(t) + G(t)K'(t)]x + G(t)u \tag{3.4-7}$$

[obtained from (3.4-1) by linear state-variable feedback] is completely controllable at t_1.

Proof. With $\bar{x}(t_1)$ and $\bar{u}(\cdot)$ as defined in the statement of the lemma, let $\bar{x}(t)$ be the resulting state of (3.4-1) at time t. Then

$$\begin{aligned}\dot{\bar{x}}(t) &= F(t)\bar{x}(t) + G(t)\bar{u}(t)\\ &= [F(t) + G(t)K'(t)]\bar{x}(t) + G(t)[\bar{u}(t) - K'(t)\bar{x}(t)],\end{aligned}$$

which is to say that if the control $\bar{u}(t) - K'(t)\bar{x}(t)$ is applied to (3.4-7) with state $\bar{x}(t_1)$ at time t_1, Eq. (3.4-7) follows the same trajectory as (3.4-1), and, in particular, reaches the zero state at time t_2. Equivalently, (3.4-7) is completely controllable at time t_1 and the lemma is established.

The control law for the infinite-time problem will be constant if F, G, Q, R, and S are all constant, as may easily be seen. The closed-loop system will be asymptotically stable if the pair $[F - GR^{-1}S', D]$ is completely observable, where D is any matrix such that $DD' = Q - SR^{-1}S'$. (If D were such that $DD' = Q$, note that complete observability of $[F, D]$ or $[F - GR^{-1}S', D]$ would not necessarily imply asymptotic stability.)

Problem 3.4-1. Generalize the lemma of this section to nonlinear systems of the form $\dot{x} = f(x, u)$ with nonlinear feedback.

Problem 3.4-2. Can you suggest a physical situation where the optimization problem associated with (3.4-1) and (3.4-2) would be appropriate?

Problem 3.4-3. State and prove a result dual to that of the lemma concerning the complete observability of a class of systems.

REFERENCES

[1] Sannuti, P., and P. V. Kokotovic, "Near Optimum Design of Linear Systems by a Singular Perturbation Method," *IEEE Trans. Auto. Control*, Vol. AC–14, No. 1, February 1969, pp. 15–21.

[2] Brkic, M., and P. V. Kokotovic, "Synthesis of a Voltage Regulator for Large A.C. Generators by Mitrovic's Method," *Proc. 6th Yugoslav Nat. ETAN Conf.* Belgrade, 1961.

[3] Larson, R. E., and R. M. Dressler, "Optimal Control of a Class of Partially Controlled Systems," *Proc. Hawaii Intern. Conf. System Sci.*, 1968, pp. 23–26.

[4] Kalman, R. E., and J. E. Bertram, "Control System Analysis and Design by the Second Method of Lyapunov," *Trans. ASME Ser. D: J. Basic Eng.* Vol. 82, No. 2, June 1960, pp. 371–400.

[5] LaSalle, J., and S. Lefschetz, "*Stability by Lyapunov's Direct Method with Applications*," Academic Press Inc., New York, 1961.

[6] Kalman, R. E., "Contributions to the Theory of Optimal Control," *Bol. Soc. Matem. Mex.*, 1960, pp. 102–119.

[7] Kalman, R. E., "When is a Linear Control System Optimal?" *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 86, March 1964, pp. 1–10.

[8] Gantmacher, F. R., *The Theory of Matrices*, Chelsea Publishing Co., New York, 1959.

CHAPTER 4

THE REGULATOR WITH A PRESCRIBED DEGREE OF STABILITY

4.1 QUALITATIVE STATEMENT OF THE REGULATOR PROBLEM WITH A PRESCRIBED DEGREE OF STABILITY

Throughout this section, we shall normally restrict attention to time-invariant systems (F and G constant) of the form

$$\dot{x} = Fx + Gu. \tag{4.1-1}$$

Although the results presented in this chapter extend to time-varying systems, they prove to be more meaningful for time-invariant systems.

We shall also be interested in considering the effect of applying control laws of the form

$$u = K'x \tag{4.1-2}$$

to (4.1-1), where K is a constant matrix. The associated closed-loop system is, of course,

$$\dot{x} = (F + GK')x. \tag{4.1-3}$$

So far, we have described one procedure for the selection of control laws, this being based on the formulation of an infinite-time optimization problem. In fact, we recall from Sec. 3.3 that if (1) $[F, G]$ is completely controllable, and (2) Q is symmetric nonnegative definite and R is symmetric

positive definite, where Q and R are both constant, then the problem of minimizing the performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u'Ru + x'Qx)\, dt \tag{4.1-4}$$

is solved by using a feedback law of the form of (4.1-2). (For the moment, we shall not be concerned with the scheme for calculating K.) We recall, too, from Sec. 3.3 that if (3) $[F, D]$ is completely observable, where D is any matrix such that $DD' = Q$, then the resulting closed-loop system is asymptotically stable.

There are, of course, other procedures for determining control laws of the form (4.1-2), which might achieve goals other than the minimization of a performance index such as (4.1-4). One other goal is to seek to have all the eigenvalues of the closed-loop system (4.1-3) taking prescribed values. The task of choosing an appropriate K has been termed the pole-positioning problem.

To solve the pole-positioning problem, the complete controllability of $[F, G]$ is required. (For a proof of this result, see [1].) The task of computing K in the single-input case is actually quite straightforward, if one converts F to companion matrix form and g to a vector containing all zeros except in the last position. (The complete controllability of $[F, g]$ is actually sufficient to ensure the existence of a basis transformation in the state space taking F and g to the special form; see Appendix B.) Now if

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_1 & -a_2 & -a_3 & \cdots & -a_n \end{bmatrix} \qquad g = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix},$$

a choice of $k' = [k_1 \quad k_2 \quad \cdots \quad k_n]$ causes

$$F + gk' = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_1 + k_1 & -a_2 + k_2 & -a_3 + k_3 & \cdots & -a_n + k_n \end{bmatrix}$$

and the eigenvalues of $F + gk'$ become the roots of the equation

$$s^n + (a_n - k_n)s^{n-1} + \cdots + (a_2 - k_2)s + (a_1 - k_1) = 0.$$

If these eigenvalues are prescribed, the determination of the feedback vector k, knowing a_1 through a_n, is immediate. For the multiple-input case, where G is no longer a vector but a matrix, the computational task is much harder but is nevertheless possible (see [2] through [6]).

From the practical point of view, the essential problem may not be to fix precisely the eigenvalues of $F + GK'$, but rather to ensure that these eigenvalues be within a certain region of the complex plane. Typical regions might be that sector of the left half-plane $Re[s] < 0$ bounded by straight lines extending from the origin and making angles $\pm\theta$ with the negative real axis, or, again, they might be that part of the left half-plane to the left of $Re[s] = -\alpha$, for some $\alpha > 0$.

We shall be concerned in this chapter with achieving the latter restriction. Moreover, we shall attempt to achieve the restriction not by selecting K through some modification of the procedure used for solving the pole-positioning problem, but by posing a suitable version of the regulator problem. Essentially what we are after is a solution of the regulator problem that gives a constant control law and that gives not merely an asymptotically stable closed-loop system, but one with a degree of stability of at least α. In other words, nonzero initial states of the closed-loop system (4.1-3) should decay at least as fast as $e^{-\alpha t}$. This is equivalent to requiring the eigenvalues of $F + GK'$ to have real parts less than $-\alpha$.

As we shall illustrate, commencing in the next section, it proves appropriate to replace the performance index (4.1-4) usually employed in the time-invariant regulator problem by a performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} e^{2\alpha t}(u'Ru + x'Qx)\, dt \tag{4.1-5}$$

where R and Q are as before—i.e., constant, symmetric, and respectively positive and nonnegative definite. The constant α is nonnegative (of course, $\alpha = 0$ corresponds to the situation considered in Chapter 3).

Since the integrand in (4.1-5) can be rewritten as $u'\hat{R}u + x'\hat{Q}x$, where $\hat{R} = Re^{2\alpha t}$, $\hat{Q} = Qe^{2\alpha t}$, the results summarized in Sec. 3.3 will guarantee that a linear feedback law will generate the optimal control for (4.1-5). That this law should be constant is not at all clear, but it will be proved in the next section. However, supposing that this is proven, we can give plausible reasons as to why the closed-loop system should have a degree of stability of at least α. It would be expected that the optimal value of (4.1-5) should be finite. Now, the integrand will be of the form $e^{2\alpha t}x'(t)Mx(t)$ for some constant M when u is written as a linear constant function of x. For the integrand to approach zero as t approaches infinity, it is clearly sufficient, and probably necessary, for $x(t)$ to decay faster than $e^{-\alpha t}$. This is equivalent to requiring the closed-loop system to have a degree of stability of at least α. The results of this chapter first appeared in [7].

4.2 QUANTITATIVE STATEMENT AND SOLUTION OF THE REGULATOR PROBLEM WITH A PRESCRIBED DEGREE OF STABILITY

The remarks of the previous section justify the formal statement of the modified regulator problem in the following manner.

Modified regulator problem. Consider the system

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{4.2-1}$$

where F and G are constant and the pair $[F, G]$ is completely controllable. Consider also the associated performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} e^{2\alpha t}(u'Ru + x'Qx)\, dt \tag{4.2-2}$$

where R and Q are constant, symmetric, and respectively positive definite and nonnegative definite. Let α be a nonnegative constant (which will turn out to be the minimum degree of stability of the closed-loop system). With D any matrix such that $DD' = Q$, let $[F, D]$ be completely observable. Define the minimization problem as the task of finding the minimum value of the performance index (4.2-2) and the associated optimal control.

The reason for the complete controllability condition is the same as explained in the last chapter; it ensures that the infinite, as distinct from finite, time problem has a solution. The observability condition will be needed to establish the constraint on the degree of stability of the closed-loop system.

The strategy we adopt in solving this modified problem is to introduce transformations that convert the problem to an infinite-time regulator problem of the type considered in the last chapter. Accordingly, we make the definitions

$$\hat{x}(t) = e^{\alpha t}x(t) \tag{4.2-3}$$

$$\hat{u}(t) = e^{\alpha t}u(t). \tag{4.2-4}$$

Just as $x(\cdot)$ and $u(\cdot)$ may be related [via Eq. (4.2-1)], so $\hat{x}(\cdot)$ and $\hat{u}(\cdot)$ may be related. Observe that

$$\begin{aligned}\dot{\hat{x}} &= \frac{d}{dt}(e^{\alpha t}x(t)) = \alpha e^{\alpha t}x(t) + e^{\alpha t}\dot{x}(t)\\ &= \alpha\hat{x} + e^{\alpha t}Fx + e^{\alpha t}Gu\\ &= (F + \alpha I)\hat{x} + G\hat{u}. \end{aligned}\tag{4.2-5}$$

Thus, given the relations (4.2-3) and (4.2-4), the system equation (4.2-1)

implies the system equation (4.2-5). The converse is clearly true, too. Corresponding initial conditions for the two systems (4.2-1) and (4.2-5) are given by setting $t = t_0$ in (4.2-3)—i.e., $\hat{x}(t_0) = e^{\alpha t_0}x(t_0)$.

The integrand in (4.2-2) may also be written in terms of $\hat{u}$ and $\hat{x}$:

$$e^{2\alpha t}(u'Ru + x'Qx) = \hat{u}'R\hat{u} + \hat{x}'Q\hat{x}.$$

Consequently, we may associate with the system (4.2-5) the performance index

$$\hat{V}(\hat{x}(t_0), \hat{u}(\cdot), t_0) = \int_{t_0}^{\infty} (\hat{u}'R\hat{u} + \hat{x}'Q\hat{x})\, dt. \tag{4.2-6}$$

Moreover, there is a strong connection between the minimization problem associated with the equation pair (4.2-1), (4.2-2), and the pair (4.2-5), (4.2-6). Suppose $u^*(t)$ is the value of the optimal control at time t for the first problem, and that $x(t)$ is the resulting value of the state at time t when the initial state is $x(t_0)$. Then the value of the optimal control at time t for the second problem is $\hat{u}^*(t) = e^{\alpha t}u^*(t)$, and the resulting value of the state at time t is given by $\hat{x}(t) = e^{\alpha t}x(t)$, provided $\hat{x}(t_0) = e^{\alpha t_0}x(t_0)$. Moreover, the minimum performance index is the same for each problem.

Moreover, if the optimal control for the second problem can be expressed in feedback form as

$$\hat{u}^*(t) = k(\hat{x}(t), t), \tag{4.2-7}$$

then the optimal control for the first problem may also be expressed in feedback form; thus,

$$u^*(t) = e^{-\alpha t}\hat{u}^*(t) = e^{-\alpha t}k(e^{\alpha t}x(t), t). \tag{4.2-8}$$

[We know that the control law (4.2-7) should be a linear one; and, indeed, we shall shortly note the specific law; the point to observe here is that a *feedback* control law for the second problem readily yields a *feedback* control law for the first problem, irrespective of the notion of linearity.]

Our temporary task is now to study the system (4.2-5), repeated for convenience as

$$\dot{\hat{x}} = (F + \alpha I)\hat{x} + G\hat{u} \qquad \hat{x}(t_0) \text{ given} \tag{4.2-5}$$

and to select a control $\hat{u}^*(\cdot)$ that minimizes the performance index

$$\hat{V}(\hat{x}(t_0), \hat{u}(\cdot), t_0) = \int_{t_0}^{\infty} (\hat{u}'R\hat{u} + \hat{x}'Q\hat{x})\, dt \tag{4.2-6}$$

where R and Q have the constraints imposed at the beginning of the section.

As we discussed in the last chapter, this minimization problem may not have a solution without additional constraints of stability or controllability. One constraint that will guarantee existence of an optimal control is, however, a requirement that $[F + \alpha I, G]$ be completely controllable. As will be shown, *this is implied by the restriction that* $[F, G]$ *is completely con-*

trollable, which was imposed in our original statement of the modified regulator problem. To see this, we first need to observe a property of the complete controllability concept, derivable from the definition given earlier. This property, also given in Appendix B, is that $[F, G]$ is completely controllable if and only if the equation $w'e^{Ft}G = 0$ for all t where w is a constant vector implies $w = 0$.

The complete controllability of $[F + \alpha I, G]$ follows from that of $[F, G]$ (and vice versa) by observing the equivalence of the following four statements.

1. $[F, G]$ is completely controllable.
2. For constant w and all t, $w'e^{Ft}G = 0$ implies $w = 0$.
3. For constant w and all t, $w'e^{\alpha t}e^{Ft}G = w'e^{(F+\alpha I)t}G = 0$ implies $w = 0$.
4. $[F + \alpha I, G]$ is completely controllable.

Given this complete controllability constraint, we can define a solution to the preceding minimization problem with the most minor of modifications to the material of Sec. 3.3. Let $P(t, T)$ be the solution at time t of the equation

$$-\dot{P} = P(F + \alpha I) + (F' + \alpha I)P - PGR^{-1}G'P + Q \tag{4.2-9}$$

with boundary condition $P(T, T) = 0$. Define

$$\bar{P} = \lim_{t \to -\infty} P(t, T), \tag{4.2-10}$$

which is a constant matrix, satisfying the steady state version of (4.2-9):

$$\bar{P}(F + \alpha I) + (F' + \alpha I)\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q = 0. \tag{4.2-11}$$

Then the optimal control becomes

$$\hat{u}^*(t) = -R^{-1}G'\bar{P}\hat{x}(t). \tag{4.2-12}$$

It is interesting to know whether the application of the control law (4.2-12) to the open-loop system (4.2-5) results in an asymptotically stable closed-loop system. We recall from the results of the last chapter that a sufficient condition ensuring this asymptotic stability is that $[F + \alpha I, D]$ should be completely observable, where D is any matrix such that $DD' = Q$. Now, just as the complete controllability of $[F, G]$ implies the complete controllability of $[F + \alpha I, G]$, so by duality, the complete observability of $[F, D]$ implies the complete observability of $[F + \alpha I, D]$. Since the complete observability of $[F, D]$ was assumed in our original statement of the regulator problem with degree of stability constraint, it follows that $[F + \alpha I, D]$ is completely observable and that the closed-loop system

$$\dot{\hat{x}} = (F + \alpha I - GR^{-1}G'\bar{P})\hat{x} \tag{4.2-13}$$

is asymptotically stable.

We can now apply these results to the original optimization problem. Recall that we need to demonstrate, first, that the optimal control law is a constant, linear feedback law, and second, that the degree of stability of the

closed-loop system is at least α. Furthermore, we need to find the minimum value of (4.2-2).

Equations (4.2-7) and (4.2-8) show us that

$$u^*(t) = -e^{-\alpha t}R^{-1}G'\bar{P}e^{\alpha t}x(t) = -R^{-1}G'\bar{P}x(t). \tag{4.2-14}$$

This is the desired constant control law; note that it has the same structure as the control law of (4.2-12).

To demonstrate the degree of stability, we have from (4.2-3) that $x(t) = e^{-\alpha t}\hat{x}(t)$. Since the closed-loop system (4.2-13) has been proved asymptotically stable, we know that $\hat{x}(t)$ approaches zero as t approaches infinity, and, consequently, that $x(t)$ approaches zero at least as fast as $e^{-\alpha t}$ when t approaches infinity.

The minimum value achieved by (4.2-2) is the same as the minimum value achieved by (4.2-6). As was shown in the previous chapter, the optimal performance index for (4.2-6) is expressible in terms of $\bar{P}$ as $\hat{x}'(t_0)\bar{P}\hat{x}(t_0)$. Consequently, the minimum value achieved by (4.2-2) is $x'(t_0)e^{-2\alpha t_0}\bar{P}x(t_0)$. Let us now summarize the results in terms of the notation used in this chapter.

> ***Solution of the regulator problem with prescribed degree of stability.*** The optimal performance index for the modified regulator problem stated at the start of this section is $x'(t_0)e^{-2\alpha t_0}\bar{P}x(t_0)$, where $\bar{P}$ is defined as the limiting solution of the Riccati equation (4.2-9) with boundary condition $P(T, T) = 0$. The matrix $\bar{P}$ also satisfies the algebraic equation (4.2-11). The associated optimal control is given by the constant linear feedback law (4.2-14), and the closed-loop system has degree of stability of at least α.

One might well ask if it is possible to construct a performance index with α equal to zero such that the control law resulting from the minimization is the same as that obtained from the preceding problem when α is nonzero. The answer is yes (see Problem 4.2-2). In other words, there are sets of pairs of matrices R and Q such that the associated regulator problem with zero α leads to a closed-loop system with degree of stability α. However, it does not appear possible to give an explicit formula for writing down these matrices without first solving a regulator problem with α nonzero.

By way of example, consider an idealized angular position control system where the position of the rotation shaft is controlled by the torque applied, with no friction in the system. The equation of motion is

$$J\ddot{\theta} = T$$

where θ is the angular position, T is the applied torque, and J is the moment of inertia of the rotating parts. In state-space form, this becomes

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ a \end{bmatrix} u$$

where $x_1 = \theta$, $x_2 = \dot{\theta}$, and $a = T/J$. As a performance index guaranteeing a degree of stability $\alpha = 1$, we choose

$$\int_0^\infty e^{2t}(u^2 + x_1^2)\, dt.$$

The appropriate algebraic equation satisfied by $\bar{P}$ is

$$\bar{P}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\bar{P} - \bar{P}\begin{bmatrix} 0 & 0 \\ 0 & a^2 \end{bmatrix}\bar{P} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = 0.$$

One possible way to solve this equation is to write down simultaneous equations for the entries of $\bar{P}$; thus,

$$\begin{aligned} 2\bar{p}_{11} - \bar{p}_{12}^2 a^2 + 1 &= 0 \\ \bar{p}_{11} + 2\bar{p}_{12} - \bar{p}_{12}\bar{p}_{22}a^2 &= 0 \\ 2\bar{p}_{12} + 2\bar{p}_{22} - \bar{p}_{22}^2 a^2 &= 0. \end{aligned}$$

These equations lead to

$$\begin{aligned} \bar{p}_{11} &= \frac{1}{a^2}\left[2 + 2\sqrt{1 + a^2} + \frac{1}{2}(2 + 2\sqrt{1 + a^2})^{3/2}\right] \\ \bar{p}_{12} &= \frac{1}{a^2}[1 + \sqrt{1 + a^2} + \sqrt{2 + 2\sqrt{1 + a^2}}] \\ \bar{p}_{22} &= \frac{1}{a^2}[2 + \sqrt{2 + 2\sqrt{1 + a^2}}]. \end{aligned}$$

The optimal control law is

$$\begin{aligned} u^* &= -g'\bar{P}x \\ &= \begin{bmatrix} -\dfrac{1}{a}(1 + \sqrt{1 + a^2} + \sqrt{2 + 2\sqrt{1 + a^2}}) \\ -\dfrac{1}{a}(2 + \sqrt{2 + 2\sqrt{1 + a^2}}) \end{bmatrix}' \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned}$$

This is implementable with proportional plus derivative (in this case, tacho, or angular velocity) feedback. The closed-loop system equation is

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -(1 + \sqrt{1 + a^2} + \sqrt{2 + 2\sqrt{1 + a^2}}) & -(2 + \sqrt{2 + 2\sqrt{1 + a^2}}) \end{bmatrix} x$$

for which the characteristic polynomial is

$$s^2 + (2 + \sqrt{2 + 2\sqrt{1 + a^2}})s + (1 + \sqrt{1 + a^2} + \sqrt{2 + 2\sqrt{1 + a^2}}).$$

It is readily checked that the roots of this polynomial are complex for all a;

therefore, the real part of the closed-loop poles is

$$-1 - \tfrac{1}{2}\sqrt{2 + 2\sqrt{1 + a^2}} < -1.$$

Thus, the requisite degree of stability is achieved.

Problem 4.2-1. Consider the system (with constant F and G)

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given}$$

and the associated performance index

$$\int_{t_0}^{\infty} e^{2\alpha t}[(u'u)^5 + (x'Qx)^5]\, dt$$

where Q is a constant nonnegative definite matrix. Find a related linear system and performance index where the integrand in the performance index is not specifically dependent on time (although, of course, it depends on u and x). Show that if an optimal feedback law exists for this related system and performance index, it is a constant law, and that from it an optimal feedback law may be determined for the original system and performance index.

Problem 4.2-2. Consider the system

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given}$$

where F and G are constant and $[F, G]$ is completely controllable. Show that associated with any performance index of the form

$$\int_{t_0}^{\infty} e^{2\alpha t}(u'Ru + x'Qx)\, dt,$$

where R is constant and positive definite, Q is constant and nonnegative definite, and α is positive, there is a performance index

$$\int_{t_0}^{\infty} (u'Ru + x'\hat{Q}x)\, dt,$$

where $\hat{Q}$ is constant and nonnegative definite, such that the optimal controls associated with minimizing these indices are the same. [*Hint:* Define $\hat{Q}$ using the solution of the first minimization problem.]

Problem 4.2-3. Consider the system

$$\dot{x} = ax + u$$

with performance index

$$\int_{t_0}^{\infty} e^{2\alpha t}(u^2 + bx^2)\, dt$$

where a, b, and α are constants, with b and α nonnegative. Examine the eigenvalue of the closed-loop system matrix obtained by implementing an optimal feedback control, and indicate graphically the variation of this eigenvalue as one or more of a, b, and α vary.

Problem 4.2-4. Consider the modified regulator problem as stated at the beginning of the section, and let $\bar{P}$ be defined as in Eqs. (4.2-9) through (4.2-11).

Show that $V(x) = x'(t)\bar{P}x(t)$ is a Lyapunov function for the closed-loop system with the property that $\dot{V}/V \leq -2\alpha$. (This constitutes another proof of the degree of stability property.)

Problem 4.2-5. Suppose that you are given a linear time-invariant system with a feedback law K_α derived from the minimization of a performance index

$$\int_{t_0}^{\infty} e^{2\alpha t}(u'u + x'x)\, dt.$$

Suppose also that experiments are performed to determine the transient response of the system for the following three cases.

1. $K_\alpha = K_0$—i.e., α is chosen as zero.
2. $K_\alpha = K_{\alpha_1}$—i.e., α is chosen as a "large," positive constant.
3. $K_\alpha = K_0 + (K_{\alpha_1} - K_0)(1 - e^{-t})$.

Give sketches of the possible transient responses.

Problem 4.2-6. If F and G are constant matrices and F is $n \times n$, it is known that $[F, G]$ is completely controllable if and only if the matrix $[G\ FG \cdots F^{n-1}G]$ has rank n. Prove that complete controllability of $[F, G]$ implies complete controllability of $[F + \alpha I, G]$ by showing that if the matrix $[G\ FG \cdots F^{n-1}G]$ has rank n, so, too, does the matrix $[G\ (F + \alpha I)\, G \cdots (F + \alpha I)^{n-1}G]$.

Problem 4.2-7. Imagine two optimization problems of the type considered in this chapter with the same F, G, Q, and R but with two different α—viz., α_1 and α_2, with $\alpha_1 > \alpha_2$. Show that $\bar{P}_{\alpha_1} - \bar{P}_{\alpha_2}$ is positive definite.

4.3 EXTENSIONS OF THE PRECEDING RESULTS

We now ask if there are positive functions $f(t)$ other than $e^{\alpha t}$ with the property that the minimization of

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} f(t)(u'Ru + x'Qx)\, dt, \tag{4.3-1}$$

given that

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{4.3-2}$$

leads to a linear constant control law. (Here, the usual constraints on F, G, Q, and R, including constancy, are assumed to apply.) Reference [8] suggests that maybe $f(t) = t^k$ could lead to a constant control law.

We show here that essentially the only possible $f(t)$ are those we have already considered—i.e., those of the form $e^{\alpha t}$—but with the constraint that α should simply be real (rather than nonnegative, as required earlier in this chapter).

By writing the integrand of (4.3-1) as $u'[f(t)R]u + x'[f(t)Q]x$, it is evident

that the associated optimal control is

$$u^*(t) = -f^{-1}(t)R^{-1}G'\bar{P}(t)x(t) \tag{4.3-3}$$

where $\bar{P}(\cdot)$ is a solution of

$$-\dot{\bar{P}} = \bar{P}F + F'\bar{P} - f^{-1}(t)\bar{P}GR^{-1}G'\bar{P} + f(t)Q. \tag{4.3-4}$$

Now (4.3-3) is to be a constant control law. Hence, the matrix $\bar{P}(t)G$ must be of the form $f(t)M$, where M is some constant matrix. Premultiply (4.3-4) by G' and postmultiply by G to obtain

$$-G'M\dot{f} = [M'FG + G'F'M - G'MR^{-1}M'G + G'QG]f. \tag{4.3-5}$$

Suppose $G'\bar{P}(t)G = G'M$ is nonzero. Then there is at least one entry—say, the i-j entry—that is nonzero. Equating the i-j entries on both sides of (4.3-5), we obtain

$$\dot{f} = \alpha f$$

for some constant α. Thus, the claim that f has the form $e^{\alpha t}$ is established for the case when $G'\bar{P}(t)G \neq 0$ for some t. Now suppose $G'\bar{P}(t)G \equiv 0$. Then, since $\bar{P}(t)$ is nonnegative definite, it follows that $\bar{P}(t)G \equiv 0$, and thus the optimal control is identically zero. In this clearly exceptional case, it is actually possible to tolerate $f(t)$ differing from $e^{\alpha t}$. Thus, we have the trivial example of

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} f(t)u'Ru\, dt,$$

which for arbitrary positive $f(t)$ and arbitrary F, G, and R has the constant optimal feedback law $u \equiv 0$.

Since the preceding two sections have restricted attention to the case of nonnegative α, it might well be asked why this restriction was made. The answer is straightforward: Constant control laws certainly result from negative α, but the resulting closed-loop systems may be unstable, although no state grows faster than $e^{-\alpha t}$. (Note that $e^{-\alpha t}$ is a growing exponential, because α is negative.)

An open problem for which a solution would be of some interest (if one exists) is to set up a regulator problem for an arbitrary system (4.3-2) such that the eigenvalues of the closed-loop system matrix $F - GR^{-1}G'\bar{P}$ possess a prescribed *relative stability*—i.e., the eigenvalues have negative real part—and if written as $\sigma + j\omega$ (σ, ω real, $j = \sqrt{-1}$), the constraint $|\omega|/|\sigma| < k$ for some prescribed constant k is satisfied.

A second open problem is to set up a regulator problem such that the closed-loop system matrix possesses a dominant pair of eigenvalues (all but one pair of eigenvalues should have a large negative real part). Partial solutions to this problem will be presented in Chapter 5.

REFERENCES

[1] Popov, V. M., "Hyperstability and Optimality of Automatic Systems with Several Control Functions," *Rev. Roumaine Sci. Tech.—Electrotechnol. Energ.*, Vol. 9, No. 4, 1964, pp. 629–690.

[2] Gopinath, B., and B. O. Lange, "On the Identification and Control of Linear Systems," *Tech. Rep. SUDAAR No. 351*, June 1968, Dept. of Aeronautics and Astronautics, Stanford Univ., Stanford, Calif.

[3] Wonham, W. M., "On Pole Assignment in Multi-Input Controllable Linear Systems," *IEEE Trans. Auto. Control*, Vol. AC–12, No. 6, December 1967, pp. 660–665.

[4] Rosenbrock, H. H., "The Assignment of Closed Loop Poles," *Control Systems Centre Rep. No. 20*, October 1967, Control Systems Centre, University of Manchester, England.

[5] Singer, R. A., "The Design and Synthesis of Linear Multivariable Systems with Application to State Estimation," *Tech. Rep. 6302–8*, June 1968, Stanford Electronics Laboratories, Stanford, Calif.

[6] Anderson, B. D. O., and D. G. Luenberger, "Design of Multivariable Feedback Systems," *Proc. IEE*, Vol. 114, No. 3, March 1967, pp. 395–399.

[7] Anderson, B. D. O., and J. B. Moore, "Linear System Optimization with Prescribed Degree of Stability," *Proc. IEE*, Vol. 116, No. 12, December 1969, pp. 2083–2087.

[8] Kalman, R. E., "When is a Linear Control System Optimal?" *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 86, March 1964, pp. 1–10.

PART III

PROPERTIES AND APPLICATION OF THE OPTIMAL REGULATOR

CHAPTER 5

PROPERTIES OF REGULATOR SYSTEMS WITH A CLASSICAL CONTROL INTERPRETATION

5.1 THE REGULATOR FROM AN ENGINEERING VIEWPOINT

We have earlier intimated a desire to point out what might be termed the "engineering significance" of the regulator. Until now, we have exposed a mathematical theory for obtaining feedback laws for linear systems. These feedback laws minimize performance indices that reflect the costs of control and of having a nonzero state. In this sense, they may have engineering significance. Furthermore, we have indicated in some detail for time-invariant systems a technique whereby the closed-loop system will be asymptotically stable, and will even possess a prescribed degree of stability. This, too, has obvious engineering significance.

Again, there is engineering significance in the fact that, in distinction to most classical design procedures, the techniques are applicable to multiple-input systems, and to time-varying systems. (We have tended to avoid discussion of the latter because of the additional complexity required in, for example, assumptions guaranteeing stability of the closed-loop system. However, virtually all the results presented hitherto and those to follow are applicable in some way to this class of system.)

But there still remains a number of unanswered questions concerning the

engineering significance of the results. For example, we might well wonder to what extent it is reasonable to think in terms of state feedback when the states of a system are not directly measurable. All the preceding, and most of the following, theory is built upon the assumption that the system states are available; quite clearly, if this theory is to be justified, we shall have to indicate some technique for dealing with a situation where no direct measurement is possible. We shall discuss such techniques in a subsequent chapter. Meanwhile, we shall continue with the assumption that the system states are available.

In classical control, the notions of gain margin and phase margin play an important role. Thus, engineering system specifications will often place lower bounds on these quantities, since it has been found, essentially empirically, that if these quantities are too small system performance will be degraded in some way. For example, if for a system with a small amount of time delay a controller is designed neglecting the time delay, and if the phase margin of the closed loop is small, there may well be oscillations in the actual closed loop. The natural question now arises as to what may be said about the gain margin and phase margin (if these quantities can, in fact, be defined) of an optimal regulator.

Of course, at first glance, there can be no parallel between the dynamic feedback of the output of a system, as occurs in classical control, and the memoryless feedback of states, as in the optimal regulator. But both schemes have associated with them a closed loop. Figure 5.1-1 shows the classical

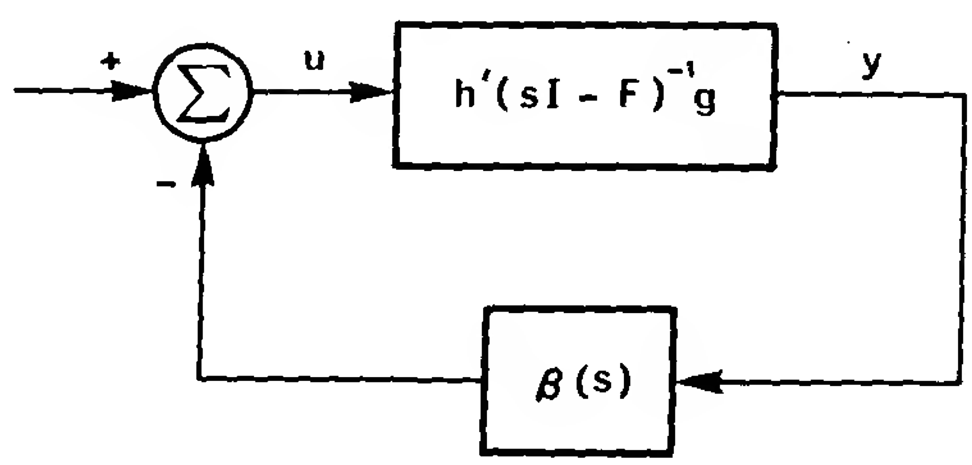


Fig. 5.1-1 Classical feedback arrangement with dynamic controller driven by system output.

feedback arrangement for a system with transfer function $h'(sI - F)^{-1}g$, where the output is fed back through a dynamic controller with transfer function $\beta(s)$. Figure 5.1-2 shows a system with transfer function $h'(sI - F)^{-1}g$ but with memoryless state-variable feedback. Here a closed loop is formed; however, it does not include the output of the open-loop system, merely the states. This closed loop is shown in Fig. 5.1-3.

Now it is clear how to give interpretations of the classical variety to the optimal feedback system. The optimal feedback system is like a classical

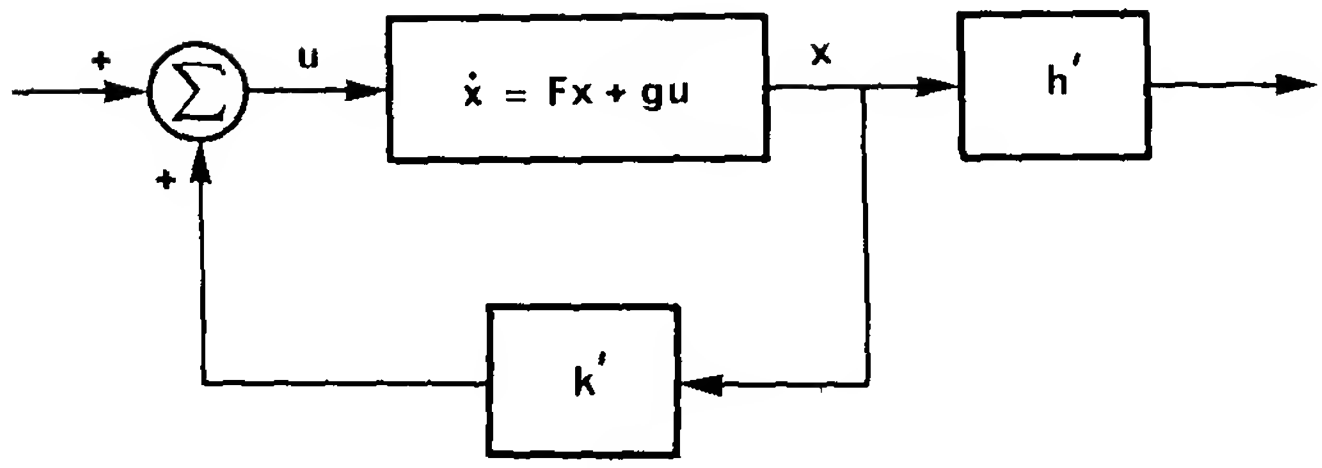


Fig. 5.1-2 Modern feedback arrangement with memoryless controller driven by system states.

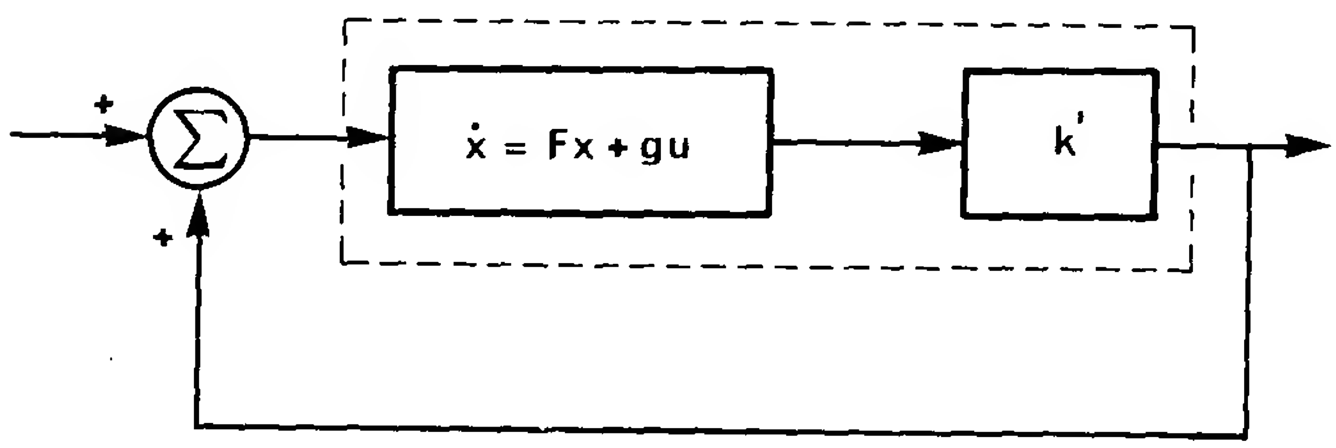


Fig. 5.1-3 The closed-loop part of a feedback system using modern feedback arrangement.

situation where unity negative feedback is applied around a (single-input, single-output) system with transfer function $-k'(sI - F)^{-1}g$. Thus, the gain margin of the optimal regulator may be determined from a Nyquist, or some other, plot of $W(j\omega) = -k'(j\omega I - F)^{-1}g$ in the usual manner.

We may recall the attention given in classical control design procedures to the question of obtaining satisfactory transient response. Thus, to obtain for a second-order system a fast response to a step input, without excessive overshoot, it is suggested that the poles of the closed-loop system should have a damping ratio of about 0.7. For a higher order system, the same sort of response can be achieved if two dominant poles of 0.7 damping ratio are used. We shall discuss how such effects can also be achieved using an optimal regulator; the key idea revolves around appropriate selection of the weighting matrices (Q and R) appearing in the performance index definition.

A common design procedure for systems containing a nonlinearity is to replace the nonlinearity by an equivalent linear element, and to design and analyze with this replacement. One then needs to know to what extent the true system performance will vary from the approximating system performance. As will be seen, a number of results involving the regulator can be obtained, giving comparative information of the sort wanted.

We shall also consider the question of incorporating relay control in otherwise optimal systems. There may often be physical and economic advantages in using relay rather than continuous control, provided stability

problems can be overcome; accordingly, it is useful to ask whether any general remarks may be made concerning the introduction of relays into optimal regulators.

A further question of engineering significance arises when we seek to discover how well an optimal regulator will perform with variations in the parameters of the forward part of the closed-loop system. One of the common aims of classical control (and particularly that specialization of classical control, feedback amplifier design) is to insert feedback so that the input-output performance of the closed-loop system becomes less sensitive to variations in the forward part of the system. In other words, one seeks to desensitize the performance to certain parameter variations.

The quantitative discussion of many of the ideas just touched upon depends on the application of one of several basic formulas, which are derived in the next section. Then we pass on to the real meat of the regulator ideas in this and the next two chapters.

5.2 SOME FUNDAMENTAL FORMULAS

To fix ideas for the remainder of this chapter, we shall restrict attention to closed-loop systems that are completely controllable, time invariant, and asymptotically stable. Thus, we shall take as our fundamental open-loop system

$$\dot{x} = Fx + Gu \tag{5.2-1}$$

with $[F, G]$ completely controllable. As the performance index, we take

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} e^{2\alpha t}(u'Ru + x'Qx)\,dt \tag{5.2-2}$$

with the usual constraints on Q and R, including that $[F, D]$ be completely observable for any D such that $DD' = Q$. At times, α will be zero. Since we shall be considering different values of α, we define P_α, rather than $\bar{P}$, as the unique positive definite solution of

$$P_\alpha(F + \alpha I) + (F' + \alpha I)P_\alpha - P_\alpha GR^{-1}G'P_\alpha + Q = 0. \tag{5.2-3}$$

The optimal control law is

$$u = K'_\alpha x = -R^{-1}G'P_\alpha x. \tag{5.2-4}$$

Using the notation $A_0 = A_\alpha|(\alpha = 0)$, we shall prove first that

$$\begin{aligned}[I - R^{1/2}K'_0(-sI - F)^{-1}GR^{-1/2}]'[I - R^{1/2}K'_0(sI - F)^{-1}GR^{-1/2}]\\ = I + R^{-1/2}G'(-sI - F')^{-1}Q(sI - F)^{-1}GR^{-1/2}.\end{aligned} \tag{5.2-5}$$

This is an identity that has appeared in a number of places (often merely in scalar form)—e.g., [1], [2], [3]. From (5.2-3), it follows that

$$P_0(sI - F) + (-sI - F')P_0 + K_0RK'_0 = Q.$$

Multiplying on the left by $R^{-1/2}G'(-sI - F')^{-1}$, and on the right by $(sI - F)^{-1}GR^{-1/2}$, yields

$$\begin{aligned} R^{-1/2}G'(-sI - F')^{-1}P_0GR^{-1/2} &+ R^{-1/2}G'P_0(sI - F)^{-1}GR^{-1/2} \\ &+ R^{-1/2}G'(-sI - F')^{-1}K_0RK_0'(sI - F)^{-1}GR^{-1/2} \\ &= R^{-1/2}G'(-sI - F')^{-1}Q(sI - F)^{-1}GR^{-1/2}. \end{aligned}$$

By adding I to each side, and observing that $P_0GR^{-1/2} = -K_0R^{1/2}$, Eq. (5.2-5) follows.

Let us now examine some variants of (5.2-5). With $s = j\omega$, Eq. (5.2-5) becomes

$$\begin{aligned} [I - R^{1/2}K_0'(-j\omega I - F)^{-1}GR^{-1/2}]'[I - R^{1/2}K_0'(j\omega I - F)^{-1}GR^{-1/2}] \\ = I + R^{-1/2}G'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}GR^{-1/2}. \end{aligned} \tag{5.2-6}$$

With * denoting the complex conjugate, the left-hand side is Hermitian, whereas the right-hand side is of the form $I + B'^*(j\omega)QB(j\omega)$ for a matrix B. If we adopt the notation $C_1 \geq C_2$ for arbitrary Hermitian matrices C_1 and C_2 to indicate that $C_1 - C_2$ is nonnegative, and if we use the fact that $B'^*QB \geq 0$, Eq. (5.2-6) implies that

$$[I - R^{1/2}K_0'(-j\omega I - F)^{-1}GR^{-1/2}]'[I - R^{1/2}K_0'(j\omega I - F)^{-1}GR^{-1/2}] \geq I. \tag{5.2-7}$$

For a single-input system (5.2-1), there is no loss of generality in taking R as unity. Equations (5.2-5), (5.2-6), and (5.2-7) then become

$$\begin{aligned} [1 - k_0'(-sI - F)^{-1}g][1 - k_0'(sI - F)^{-1}g] \\ = 1 + g'(-sI - F')^{-1}Q(sI - F)^{-1}g, \end{aligned} \tag{5.2-8}$$

or

$$|1 - k_0'(j\omega I - F)^{-1}g|^2 = 1 + g'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}g. \tag{5.2-9}$$

Therefore, we have that

$$|1 - k_0'(j\omega I - F)^{-1}g| \geq 1. \tag{5.2-10}$$

Problem 5.2-1 asks for the establishment of the result that the equality sign in (5.2-7) and (5.2-10) can hold only for isolated values of ω.

Relations (5.2-5) through (5.2-10) are all specialized to the case $\alpha = 0$. For the case $\alpha \neq 0$, there are effectively two ways of developing the corresponding relations. The first relies on observing, with the aid of (5.2-3), that P_α satisfies the same equation as P_0, save that $F + \alpha I$ replaces F. Consequently, Eqs. (5.2-5) through (5.2-10), derived from (5.2-3), will be supplanted by equations where $F + \alpha I$ replaces F, and K_α replaces K_0. For example, (5.2-9) yields

$$\begin{aligned} |1 - k_0'(j\omega I - F - \alpha I)^{-1}g|^2 \\ = 1 + g'(-j\omega I - F' - \alpha I)^{-1}Q(j\omega I - F - \alpha I)^{-1}g, \end{aligned} \tag{5.2-11}$$

and, evidently, one could also regard this as coming from (5.2-9) by replacing k_0 by k_α and $\pm j\omega$ by $\pm j\omega - \alpha$ (leaving F invariant).

The second way of deriving relations corresponding to (5.2-5) through (5.2-10) for the case when α is nonzero is to observe, again with the aid of (5.2-3), that P_α satisfies the same equation as P_0, save that $Q + 2\alpha P_\alpha$ replaces Q. Then Eq. (5.2-9), for example, is replaced by

$$\begin{aligned}|1 - k'_\alpha(j\omega I - F)^{-1}g|^2 &= 1 + g'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}g \\ &\quad + 2\alpha g'(-j\omega I - F')^{-1}P_\alpha(j\omega I - F)^{-1}g.\end{aligned} \tag{5.2-12}$$

The other equations in the set (5.2-5) through (5.2-10) will also yield new relations by simply replacing K_0 by K_α and Q by $Q + 2\alpha P_\alpha$.

Problem 5.2-1. Show that inequalities (5.2-7) and (5.2-10) become equalities only for isolated values of ω.

Problem 5.2-2. Show that (5.2-9) implies

$$\begin{aligned}1 - |1 + k'_0(j\omega I - F - gk'_0)^{-1}g|^2 \\ = g'(-j\omega I - F' - k_0 g')^{-1}Q(j\omega I - F - gk'_0)^{-1}g.\end{aligned}$$

State two equivalent equations applying for $\alpha \neq 0$.

Problem 5.2-3. Derive an equation analogous to (5.2-8) when $u = k'_0 x$ is the optimal control for the system $\dot{x} = Fx + gu$ with a performance index $\int_{t_0}^{\infty} (u^2 + 2x'Su + x'Qx)\,dt$, and when $Q - SS'$ is nonnegative definite symmetric.

Problem 5.2-4. Assume that for a specific value of α—say, $\alpha = \bar{\alpha}$—the matrix $P_{\bar{\alpha}}$ has been calculated as the positive definite solution of (5.2-3). As an approximate technique for computing P_α for values of α near $\bar{\alpha}$, one could set $P_\alpha = P_{\bar{\alpha}} + (\alpha - \bar{\alpha})(\partial P_\alpha/\partial\alpha)|(\alpha = \bar{\alpha})$. Show that $\partial P_\alpha/\partial\alpha$ satisfies a linear algebraic matrix equation.

5.3 GAIN MARGIN, PHASE MARGIN AND TIME-DELAY TOLERANCE

In this section, we shall restrict attention to the regulator with a scalar input, and we shall examine certain properties of the closed-loop scheme of Fig. 5.3-1, which is a redrawn version of Fig. 5.1-3 (discussed earlier in this chapter), with the gain k replaced by k_α. We shall be especially interested in the gain margin, phase margin, and time-delay tolerance of the scheme.

We recall that the gain margin of a closed-loop system is the amount by which the loop gain can be increased until the system becomes unstable.

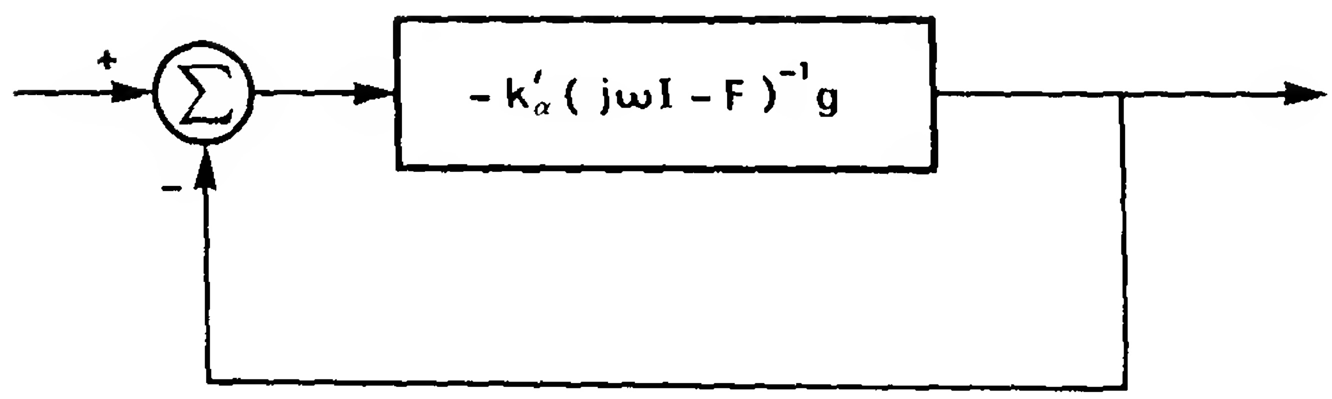


Fig. 5.3-1 Closed-loop optimal scheme with degree of stability α drawn as a unity negative feedback system.

If the loop gain can be increased without bound—i.e., instability is not encountered, no matter how large the loop gain becomes—then the closed-loop system is said to possess an infinite gain margin.

Of course, no real system has infinite gain margin. Such parasitic effects as stray capacitance, time delay, etc., will always prevent infinite gain margin from being a physical reality. Some mathematical models of systems may, however, have an infinite gain margin. Clearly, if these models are accurate representations of the physical picture—save, perhaps, for their representation of parasitic effects—it could validly be concluded that the physical system had a very large gain margin.

We shall now show that the optimally designed regulator possesses the infinite gain margin property, by noting a characteristic feature of the Nyquist diagram of the open-loop gain of the regulator. The scheme of Fig. 5.3-1 is arranged to have unity *negative* feedback, so that we may apply the Nyquist diagram ideas immediately. The associated Nyquist plot is a curve on the Argand diagram, or complex plane, obtained from the complex values of $-k'_\alpha(j\omega I - F)^{-1}g$ as ω varies through the real numbers from minus to plus infinity. Now the Nyquist plot of $-k'_\alpha(j\omega I - F)^{-1}g$ is constrained to avoid a certain region of the complex plane, because Eq. (5.2-12) of the last section yields that

$$|1 - k'_\alpha(j\omega I - F)^{-1}g| \geq 1, \tag{5.3-1}$$

which is to say that the distance of any point on the Nyquist plot from the point $-1 + j0$ is at least unity. In other words, the plot of $-k'_\alpha(j\omega I - F)^{-1}g$ avoids a circle of unit radius centered at $-1 + j0$.

See Fig. 5.3-2 for examples of various plots. (The transfer functions are irrelevant.) The arrow marks the direction of increasing ω. Note that the plots end at the origin, which is always the case when the open-loop transfer function, expressed as a numerator polynomial divided by a denominator polynomial, has the numerator degree less than the denominator degree.

There is yet a further constraint on the Nyquist plot, which is caused by the fact that the closed-loop system is known to be asymptotically stable. This restricts the number of *counterclockwise encirclements* of the point $-1 + j0$ by the Nyquist plot to being precisely the number of poles of the

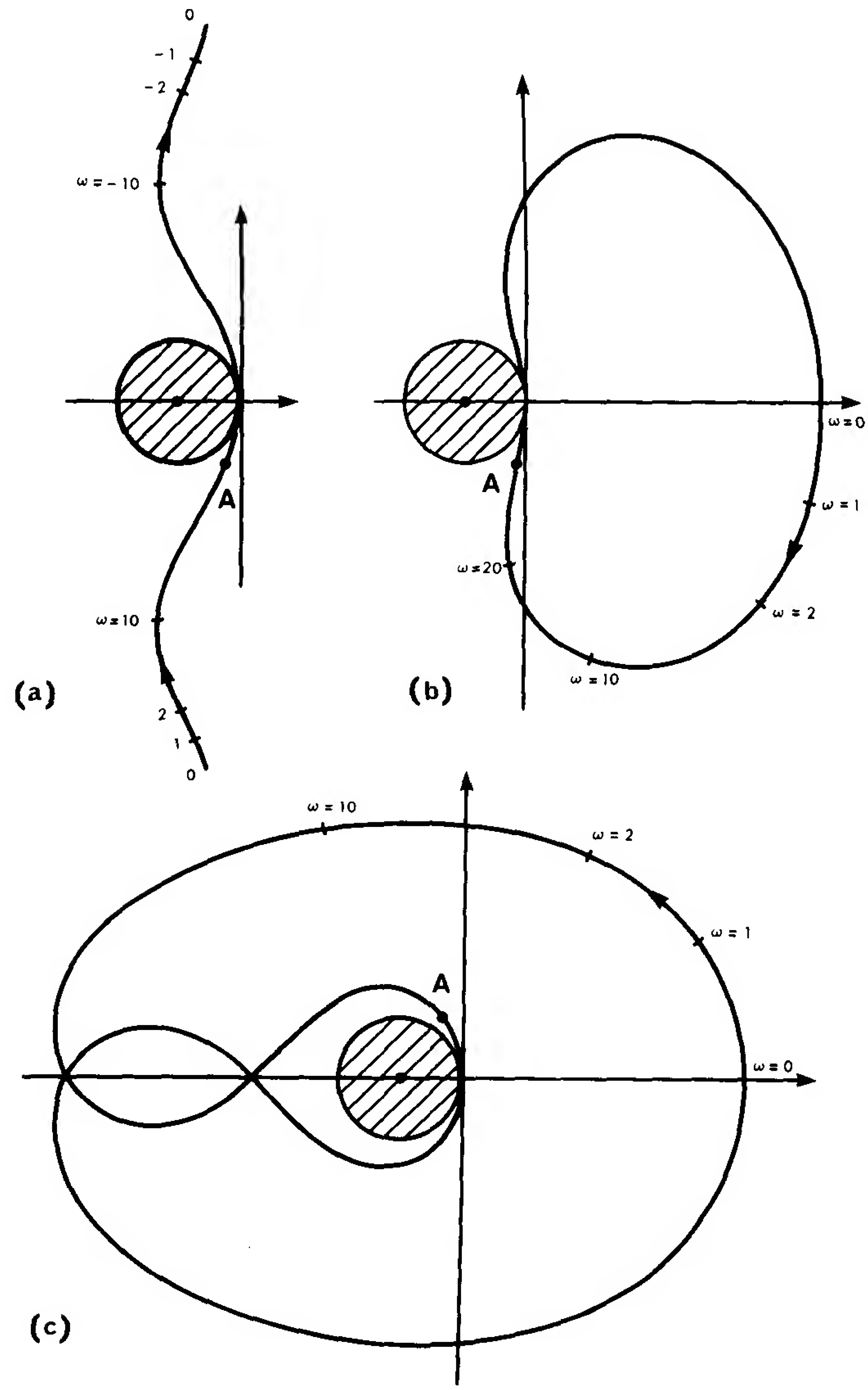


Fig. 5.3-2 Nyquist plots of $-k'_\alpha(j\omega I - F)^{-1}g$ avoiding a unit critical disc center $(-1, 0)$. Points A are at unity distance from the origin.

transfer function $-k'_\alpha(sI - F)^{-1}g$ lying in $Re[s] \geq 0$. We understand that if a pole lies on $Re[s] = 0$, the Nyquist diagram is constructed by making a small semicircular indentation into the region $Re[s] < 0$ around this pole, and plotting the complex numbers $-k'(sI - F)^{-1}g$ as s moves around this semicircular contour. (For a discussion and proof of this basic stability result, see, for example, [4].)

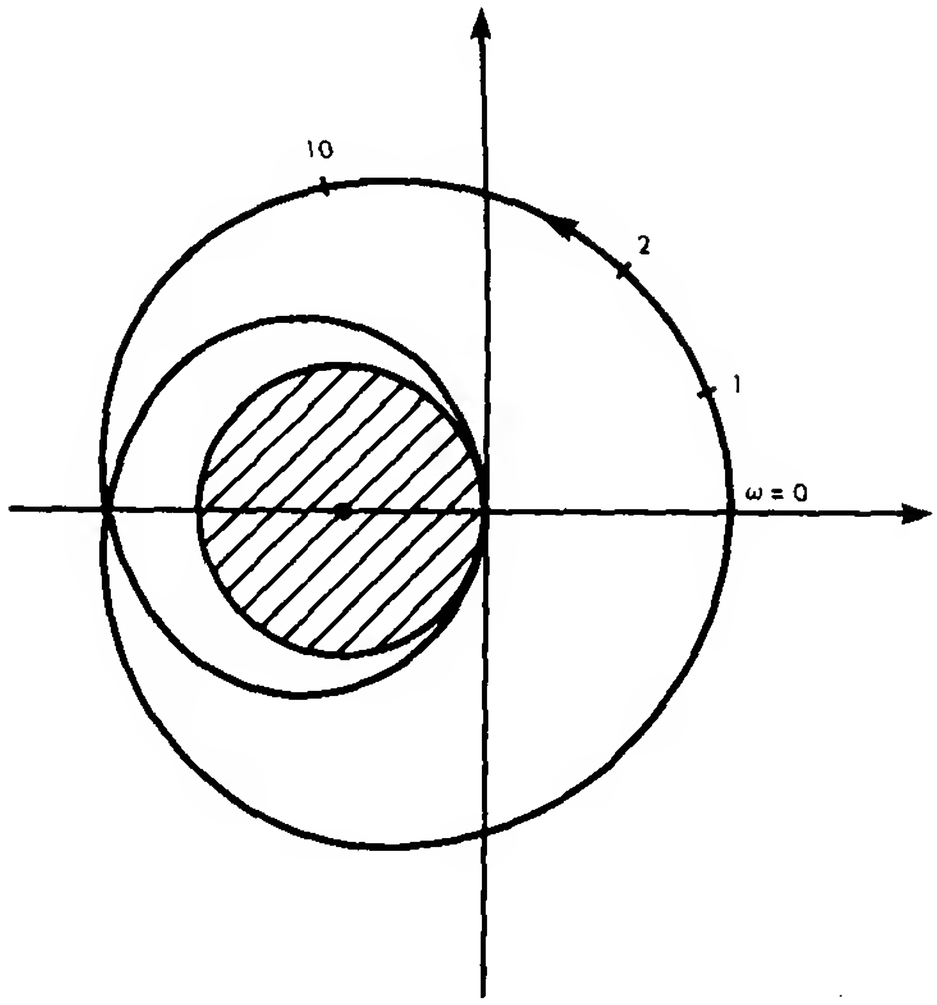


Fig. 5.3-3 Nyquist plot for which open-loop transfer function has two poles in Re[s] > 0, and closed-loop is stable.

Thus, if $-k'_\alpha(sI - F)^{-1}g$ has no poles in $Re[s] \geq 0$, diagrams such as those in Fig. 5.3-2 may be obtained. Fig. 5.3-3 illustrates a case where $-k'_\alpha(sI - F)^{-1}g$ has two poles in $Re[s] > 0$.

The key observation we now make is that *the number of encirclements of the point* $-1 + j0$ *is the same as the number of encirclements of any other point inside the circle of unit radius and center* $-1 + j0$. The best way to see this seems to be by visual inspection; a few experiments will quickly show that if the preceding remarks were not true, the Nyquist diagram would have to enter the circle from which it has been excluded.

It is known that the closed-loop system with gain multiplied by a constant factor β will continue to be asymptotically stable if the Nyquist diagram of $-\beta k'_\alpha(j\omega I - F)^{-1}g$ encircles $-1 + j0$ in a counterclockwise direction a number of times equal to the number of poles of $-\beta k'_\alpha(sI - F)^{-1}g$ lying in $Re[s] \geq 0$. Equivalently, asymptotic stability will follow if the Nyquist diagram of $-k'_\alpha(j\omega I - F)^{-1}g$ encircles $-(1/\beta) + j0$ this number of times. But our previous observation precisely guarantees this for a range of β. The points $-(1/\beta) + j0$ for all real $\beta > \frac{1}{2}$ lie inside the critical circle, and thus are encircled counterclockwise the same number of times as the point $-1 + j0$. As we have argued, this number is the same as the number of poles of $-k'_\alpha(sI - F)^{-1}g$ in $Re[s] \geq 0$. Consequently, with asymptotic stability following for all real $\beta > \frac{1}{2}$, we have established the infinite gain margin property.

Let us now turn to consideration of the phase margin property. First, we recall the definition of phase margin. It is the amount of negative phase shift that must be introduced (without gain increase) to make that part of

the Nyquist plot corresponding to $\omega \geq 0$ pass through the $-1 + j0$ point. For example, consider the three plots of Fig. 5.3-2; points A at unit distance from the origin on the $\omega \geq 0$ part of the plot have been marked. The negative phase shift that will need to be introduced in the first and second case is about 80°, and that in the third case about 280°. Thus, 80° is approximately the phase margin in the first and second case, and 280° in the third.

We shall now show that the phase margin of an optimal regulator is always at least 60°. The phase margin is determined from that point or those points on the $\omega \geq 0$ part of the Nyquist plot which are at unit distance from the origin. Since the Nyquist plot of an optimal regulator must avoid the circle with center $-1 + j0$ and unity radius, the points at unit distance from the origin and lying on the Nyquist plot of an optimal regulator are restricted to lying on the shaded part of the circle of unit radius and center the origin, shown in Fig. 5.3-4. The smallest angle through which one of the allowable

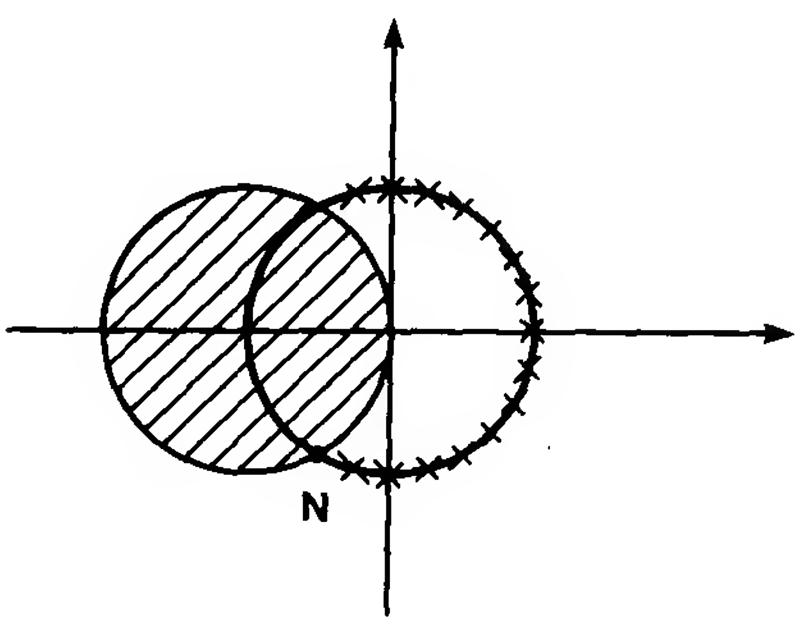


Fig. 5.3-4 Shaded points denote permissible points on Nyquist plot of optimal regulator at unit distance from origin.

points could move in a clockwise direction to reach $-1 + j0$ is 60°, corresponding to the point N of Fig. 5.3-4. Any other point in the allowed set of points (those outside the circle of center $-1 + j0$, and unity radius, but at unit distance from the origin) must move through more than 60° to reach $-1 + j0$. The angle through which a point such as N must move to reach $-1 + j0$ is precisely the phase margin. Consequently, the lower bound of 60° is established.

Let us now consider the effect of varying α away from zero. Reference to Eq. (5.2-11) shows that

$$|1 - k'_\alpha(j\omega I - F - \alpha I)^{-1}g| = |1 - k'_\alpha[(j\omega - \alpha)I - F]^{-1}g| \geq 1. \tag{5.3-2}$$

Consequently, if a modified Nyquist plot is made of $-k'_\alpha[(j\omega - \alpha)I - F]^{-1}g$ rather than of $-k'_\alpha(j\omega I - F)^{-1}g$, this modified plot will also avoid the circle of center $-1 + j0$ and radius 1. Then it will follow that if a gain $\beta > \frac{1}{2}$ is inserted in the closed loop, the degree of stability α of the closed-loop system will be retained for all real β. If a negative phase shift of up to 60° is introduced, a degree of stability α will also still be retained.

Yet another interpretation of the case of nonzero α follows by comparing Eqs. (5.2-9) and (5.2-12) of Sec. 5.2, repeated here for convenience:

$$|1 - k_0'(j\omega I - F)^{-1}g|^2 = 1 + g'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}g$$

and

$$\begin{aligned}|1 - k_\alpha'(j\omega I - F)^{-1}g|^2 = 1 &+ g'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}g \\ &+ 2\alpha g'(-j\omega I - F')^{-1}P_\alpha(j\omega I - F)^{-1}g,\end{aligned}$$

whence

$$\begin{aligned}|1 - k_\alpha'(j\omega I - F)^{-1}g|^2 = |1 &- k_0'(j\omega I - F)^{-1}g|^2 \\ &+ 2\alpha g'(-j\omega I - F')^{-1}P_\alpha(j\omega I - F)^{-1}g.\end{aligned} \tag{5.3-3}$$

The second term on the right side is nonnegative, being of the form $2\alpha b'^*(j\omega)P_\alpha b(j\omega)$ for a vector b. In fact, one can show that b is never zero, and so

$$|1 - k_\alpha'(j\omega I - F)^{-1}g| > |1 - k_0'(j\omega I - F)^{-1}g|. \tag{5.3-4}$$

This equation says that any two points on the Nyquist plots of $-k_\alpha'(j\omega I - F)^{-1}g$ and $-k_0'(j\omega I - F)^{-1}g$ corresponding to the same value of ω are such that the point on the first plot is further from $-1 + j0$ than the point on the second plot. In loose terms, the whole plot of $-k_\alpha'(j\omega I - F)^{-1}g$ is further from $-1 + j0$ than the plot of $-k_0'(j\omega I - F)^{-1}g$. This does not, however, imply that the phase margin for the former is greater than that for the latter. Problem 5.3-1 asks for a proof of this somewhat surprising result.

Problem 5.3-2 asks for a proof of the result that if $\alpha_1 > \alpha_2 > 0$, the Nyquist plot of $-k_{\alpha_1}'(j\omega I - F)^{-1}g$ is further, in the preceding sense, from $-1 + j0$ than the plot of $-k_{\alpha_2}'(j\omega I - F)^{-1}g$.

We now turn to a discussion of the tolerance of time delay in the closed loop. Accordingly, we shall consider the scheme of Figure 5.3-5, where T is a certain time delay. (The delay block could equally well be anywhere else in the loop, from the point of view of the following discussion.) We shall be concerned with the stability of the closed-loop system.

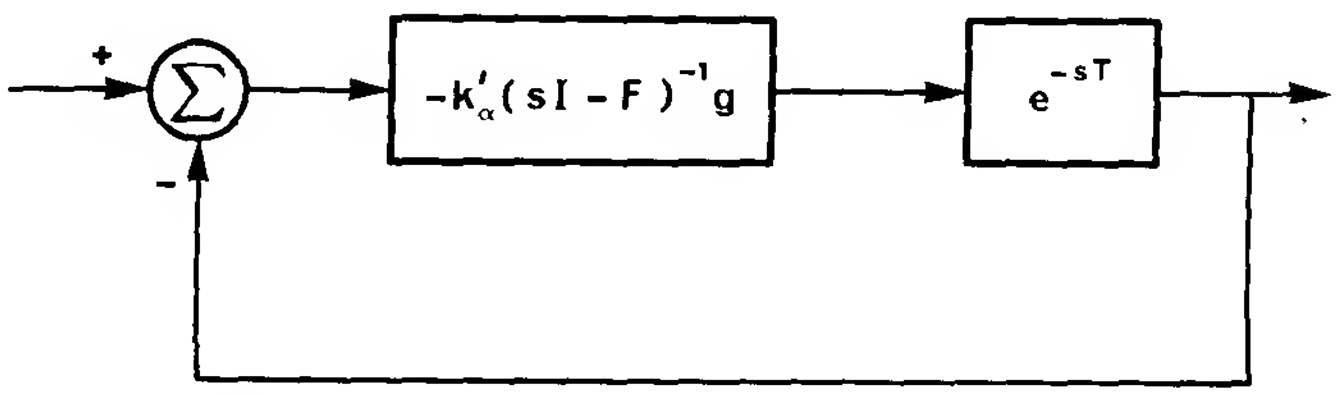


Fig. 5.3-5 Optimal regulator with time delay T inserted.

The effect of the time delay is to insert a frequency-dependent negative phase shift into the open-loop transfer function. Thus, instead of $-k'_\alpha(j\omega I - F)^{-1}g$ being the open-loop transfer function, it will be $-k'_\alpha(j\omega I - F)^{-1}ge^{-j\omega T}$. This has the same magnitude as $-k'_\alpha(j\omega I - F)^{-1}g$, but a negative phase shift of ωT radians.

It is straightforward to derive allowable values for the time delay which do not cause instability. Considering first the $\alpha = 0$ case, suppose the transfer function $-k'_0(j\omega I - F)^{-1}g$ has unity gain at the frequencies $\omega_1, \omega_2, \ldots, \omega_r$ with $0 < \omega_1 < \omega_2 < \cdots < \omega_r$, and let the amount of negative phase shift that would bring each of these unity gain points to the $-1 + j0$ point be $\phi_1, \phi_2, \ldots, \phi_r$, respectively. Of course, $\phi_i \geq \pi/3$ for all i. Then, if a time delay T is inserted, so long as $\omega_i T < \phi_i$ or $T < \phi_i/\omega_i$ for all i, stability will prevail. In particular, if $T < \pi/3\omega_r$, stability is assured.

For the case of α nonzero, since the modified Nyquist plot of $-k'_\alpha[(j\omega - \alpha)I - F]^{-1}g$ has the same properties as the Nyquist plot of $-k'_0(j\omega I - F)^{-1}g$, we see (with obvious definition of $\omega_1, \ldots, \omega_r$ and $\phi_1, \ldots, \phi_r$) that $T < \phi_i/\omega_i$ for all i, and, in particular, $T < \pi/3\omega_r$ will ensure maintenance of the degree of stability α. Greater time delays can be tolerated, but the degree of stability will be reduced.

The introduction of time delay will destroy the infinite gain margin property. To see this, observe that as ω approaches infinity, the phase shift introduced—viz., ωT radians—becomes infinitely great for any nonzero T. In particular, one can be assured that the Nyquist plot of $-k'_\alpha(j\omega I - F)^{-1}g$ will be rotated for suitably large ω such that the rotated plot crosses the real axis just to the left of the origin. (In fact, the plot will cross the axis infinitely often.) If for a given T, the leftmost point of the real axis that is crossed is $-(1/\beta) + j0$, then the gain margin of the closed loop with time delay is $20 \log_{10} \beta$ db.

Of course, the introduction of a multiplying factor, phase shift, or time delay in the loop will destroy the optimality of the original system. But the important point to note is that the optimality of the original system allows the introduction of these various modifications while maintaining the required degree of stability. Optimality may not, in any specific situation, be of direct engineering significance, but stability is; therefore, optimality becomes of engineering significance indirectly.

Problem 5.3-1. Show with suitable sketches that the relation $|1 - k'_\alpha(j\omega I - F)^{-1}g| > |1 - k'_0(j\omega I - F)^{-1}g|$ does not imply that the phase margin associated with the transfer function $-k'_\alpha(j\omega I - F)^{-1}g$ is greater than that associated with the transfer function $-k'_0(j\omega I - F)^{-1}g$.

Problem 5.3-2. Suppose $\alpha_1 > \alpha_2 > 0$. Show that $|1 - k'_{\alpha_1}(j\omega I - F)^{-1}g| > |1 - k'_{\alpha_2}(j\omega I - F)^{-1}g|$. [*Hint:* Use the result of Problem 4.2-7 of Chapter 4 that $P_{\alpha_1} - P_{\alpha_2}$ is positive definite.]

Problem 5.3-3. Suggest ways of extending the preceding material to multiple-input systems.

Problem 5.3-4. Let $u = k_0'x$ be the optimal control for the system $\dot{x} = Fx + gu$ with performance index $\int_0^\infty (u^2 + 2x'Su + x'Qx)\,dt$, where it is assumed that $Q - SS'$ is nonnegative definite symmetric. Show that it is *not* generally possible to conclude results of the sort described in this section.

5.4 POSITIONING OF THE CLOSED-LOOP SYSTEM POLES

In the last chapter, we pointed out that frequently a feedback law may be required such that the closed-loop system has certain desired poles. These poles may be specified so that, for example, a desired transient response may be obtained or a certain bandwidth may be achieved by the closed-loop system. Frequently, such specifications may call for some of the closed-loop poles to be dominant, with the implication that the real part of the other poles should be negative and large in magnitude.

In this section, we shall present an optimal design procedure that will achieve approximately the aim of fixing the dominant closed-loop poles. The end result of the procedure is a set of dominant closed-loop poles that are close to the desired dominant poles, with the remaining nondominant poles possessing a large negative real part. The system specifications that generate a requirement for dominant poles will, in general, not require the exact realization of these dominant poles. Therefore, the design procedure, although approximate, is still generally useful in terms of meeting the pole position constraints. *At the same time, an optimal system results, with all its attendant advantages.* Chapter 7, Sec. 7.4 will discuss further techniques for pole positioning in optimal systems.

A characteristic of the design procedure is that not all closed-loop poles may be dominant. Since the procedure allows specification in advance of only the dominant poles, the ability to specify all poles is lacking. What allows us to specify that the system is optimal is the lack of dominancy of precisely that pole (or those poles) which may not be specified. As remarked in the last chapter, if optimality is not desired, we can always find a linear feedback law for a linear system such that all closed-loop poles have desired values. Thus, to retain optimality, we are simply sacrificing the right to specify certain (in fact, the unimportant) pole positions.

The advantages attendant in having optimality of the closed-loop system have already been partly indicated, but they will become clearer in the next two chapters. For the moment, the reader will still have to accept the fact that optimality has more to offer than a first glance would indicate.

As preliminary material associated with the design procedure we shall also indicate the following:

1. A computational procedure (not involving the Riccati equation) for obtaining the optimal control law for a single-input time invariant system.
2. A general property of the closed-loop system poles. (This property will be exploited in defining the design procedure.)

Of course, (1) is of interest in its own right. Following discussion of (1) and (2), we shall develop the design procedure, first for single-input systems and then for multiple-input systems.

To begin with, we shall restrict consideration to single-input systems

$$\dot{x} = Fx + gu \tag{5.4-1}$$

where F is an $n \times n$ matrix, and the pair $[F, g]$ is completely controllable. We shall first derive a *computational procedure*, independent of use of the Riccati equation, for minimizing a performance index of the type

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u^2 + x'Qx)\, dt \tag{5.4-2}$$

with $Q = DD'$ for some D such that the pair $[F, D]$ is completely observable.

Under the stated controllability and observability conditions, we know that there is an optimal closed-loop law $u = k'x$, such that the closed-loop system is asymptotically stable and that the following equation [derived from the fundamental Eq. (5.2-8) of the earlier section], is satisfied:

$$\begin{aligned}[1 - k'(-sI - F)^{-1}g][1 - k'(sI - F)^{-1}g]\\ = 1 + g'(-sI - F')^{-1}Q'(sI - F)^{-1}g.\end{aligned} \tag{5.4-3}$$

Let $\psi(s)$ denote the characteristic polynomial of F—i.e., det $(sI - F)$. Define the polynomial $p(s)$ via

$$1 - k'(sI - F)^{-1}g = \frac{p(s)}{\psi(s)} \tag{5.4-4}$$

and the polynomial $q(s)$ by

$$g'(-sI - F')^{-1}Q(sI - F)^{-1}g = \frac{q(s)}{\psi(-s)\psi(s)}. \tag{5.4-5}$$

When s is set equal to infinity in this equation, the left-hand side is clearly zero. Hence, $q(s)$ has degree less than twice that of $\psi(s)$, which has degree n. On other hand, $p(s)$ has the same degree as $\psi(s)$, as may be seen by setting s equal to infinity in (5.4-4). Also, we see from (5.4-4) and the fact that $\psi(s)$ is monic that $p(s)$ is also monic—i.e., its highest order term is s^n.

The polynomial $p(s)$ has another important interpretation, which we may now derive. Observe that

$$\begin{aligned}
&[1 - k'(sI - F)^{-1}g][1 + k'(sI - F - gk')^{-1}g] \\
&\quad = 1 - k'(sI - F)^{-1}g + k'(sI - F - gk')^{-1}g \\
&\qquad - k'(sI - F)^{-1}[(sI - F) - (sI - F - gk')](sI - F - gk')^{-1}g \\
&\quad = 1.
\end{aligned}$$

Thus, inversion of (5.4-4) gives

$$\frac{\psi(s)}{p(s)} = 1 + k'(sI - F - gk')^{-1}g,$$

and, accordingly,

$$p(s) = \det(sI - F - gk'). \tag{5.4-6}$$

We recall that the closed-loop optimal system is $\dot{x} = (F + gk')x$. Therefore, the zeros of $p(s)$ are the poles of the closed-loop system transfer function. In particular, since the closed-loop system is asymptotically stable, we note that the zeros of $p(s)$ must all have negative real parts.

Equations (5.4-3), (5.4-4), and (5.4-5) combine to yield

$$\frac{p(-s)p(s)}{\psi(-s)\psi(s)} = 1 + \frac{q(s)}{\psi(-s)\psi(s)}, \tag{5.4-7}$$

which may also be written

$$p(-s)p(s) = \psi(-s)\psi(s) + q(s). \tag{5.4-8}$$

It is this equation that may be used as a basis for determining the optimal control law k, when F, g, and Q are known: All quantities on the right side are known, and thus $p(-s)p(s)$ is known. The requirement that $p(s)$ possess zeros all with negative real parts and be monic, then specifies $p(s)$ uniquely from $p(-s)p(s)$. Finally, Eq. (5.4-4) yields k. A summary of the method for determining k follows.

1. Form the polynomial $\psi(s) = \det(sI - F)$ and the polynomial $q(s) = g'(-sI - F')^{-1}Q(sI - F)^{-1}g\psi(s)\psi(-s)$. Notice that $q(s) = q(-s)$.
2. Construct the right-hand side of Eq. (5.4-8), and factor the resulting polynomial. This polynomial is even, and thus has the property that if σ is a root, so is $-\sigma$.
3. Select those roots σ_i of the polynomial having negative real parts, and construct $\prod_i(s - \sigma_i)$, which is a monic polynomial with roots these σ_i.
4. Because the zeros of $p(s)$ must have negative real parts, and because $p(s)$ is monic, it follows that $p(s)$ is uniquely determined by $p(s) = \prod_{i=1}^{n}(s - \sigma_i)$. [Note also that the constraints on $p(s)$ guarantee that no σ_i in step 3 will have zero real part.]

5. Construct k so that $1 - k'(sI - F)^{-1}g = p(s)/\psi(s)$. (This is a simple task requiring no more than the solving of linear equations which determine k uniquely.)

Let us give an example of the preceding procedure. We take

$$F = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad g = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad Q = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then $\psi(s) = \det(sI - F) = s^2 + 1$, and $q(s) = 3$. [Notice that $q(s)$ is even, as required.] Consequently, $p(-s)p(s) = (s^2 + 1)(s^2 + 1) + 3 = s^4 + 2s^2 + 4$. Now observe the factorization $(s^4 + 2s^2 + 4) = (s^2 - \sqrt{2}s + 2)(s^2 + \sqrt{2}s + 2)$, from which it is clear that $p(s) = s^2 + \sqrt{2}s + 2$. The zeros of the polynomial $p(s)$ are the closed-loop system poles. Meanwhile, $1 - k'(sI - F)^{-1}g = (s^2 + \sqrt{2}s + 2)/(s^2 + 1) = 1 + (\sqrt{2}s + 1)/(s^2 + 1)$. It is easy to check that

$$(sI - F)^{-1}g = \frac{1}{s^2 + 1}\begin{bmatrix} 1 \\ s \end{bmatrix}$$

and, consequently, $k' = [-1 \quad -\sqrt{2}]$.

We stress that this particular computation procedure is only applicable to time-invariant systems (in contrast to the approach using the Riccati equation), and could be extended but with difficulty to multiple-input systems. Moreover, the procedure involves the potentially difficult step of factoring a polynomial [the right side of (5.4-8)] and extracting out the roots with negative real parts. Nevertheless, the procedure may be valuable for computational purposes, provided the system order is not high. It has additional theoretical advantages, which we shall now illuminate.

We modify the original performance index (5.4-2), and replace it by

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u^2 + \rho x'Qx)\,dt \tag{5.4-9}$$

where ρ is a positive constant. We shall examine the *effect on the closed-loop poles of making ρ very large*. Our conclusions concerning the variation will then enable us to present the design procedure to which we referred at the beginning of the section.

Equation (5.4-8) will be replaced by

$$p(-s)p(s) = \psi(-s)\psi(s) + \rho q(s). \tag{5.4-10}$$

The poles of the closed-loop system—i.e., the zeros of $p(s)$—are given by those zeros of the polynomial $\psi(-s)\psi(s) + \rho q(s)$ having negative real parts, or, equivalently, lying in the left half-plane $Re[s] < 0$. For the moment, let us disregard the problem of separating out those particular zeros of $\psi(-s)\psi(s) + \rho q(s)$ that lie in the left half-plane, and instead concentrate on examining the behavior of all the zeros as ρ approaches infinity.

It is immediately clear that any zero of $\psi(-s)\psi(s) + pq(s)$ that remains finite as ρ approaches infinity must tend toward a zero of $q(s)$. Now $q(s)$ has degree $2n'$, which is less than the degree $2n$ of $\psi(-s)\psi(s) + pq(s)$ [see the remark immediately following (5.4-5)]. Therefore, $2n'$ zeros only of $\psi(-s)\psi(s) + pq(s)$ will remain finite as ρ approaches infinity, and the remaining $2(n - n')$ must tend to infinity. Let us consider the precise manner in which they do so. Clearly, for very large values of s, the polynomial $\psi(-s)\psi(s)$ is approximated by its highest order term—viz., $(-1)^n s^{2n}$; also, for very large values of s, the polynomial $pq(s)$ is approximated by its highest order term, which is of the form $(-1)^{n'}p\alpha s^{2n'}$ for some constant α. Now (5.4-5) implies that $q(j\omega) \geq 0$ for all real ω, which, in turn, implies that α is positive. Consequently, for large ρ, those zeros of $\psi(-s)\psi(s) + pq(s)$ that tend to infinity must approximately satisfy the equation

$$(-1)^n s^{2n} + (-1)^{n'} p\alpha s^{2n'} = 0$$

with α a certain positive constant. Since the zero roots of this equation are irrelevant, the equation may be replaced by

$$s^{2(n-n')} = (-1)^{(n-n'+1)} p\alpha. \tag{5.4-11}$$

In summary, then, as ρ approaches infinity, $2n'$ zeros of $\psi(-s)\psi(s) + pq(s)$ approach the $2n'$ zeros of $q(s)$, whereas the remaining $2(n - n')$ zeros approach the zeros of Eq. (5.4-11).

The zeros of $p(s)$, or the closed-loop system poles, are the particular zeros of $\psi(-s)\psi(s) + pq(s)$ having negative real parts. Therefore, as ρ approaches infinity, n' of the closed-loop system poles approach those zeros of $q(s)$ with negative real part, whereas $(n - n')$ approach those roots of (5.4-11) with negative real part.

The latter zeros lie in a pattern on a circle of radius $(p\alpha)^{1/2(n-n')}$, which network theory terms a Butterworth configuration [1]. The phase angles are given in the following table.

$n - n' = 1$	$+180°$
$n - n' = 2$	$\pm 135°$
$n - n' = 3$	$\pm 120°$, $+180°$
$n - n' = 4$	$\pm 112.5°$, $\pm 157.5°$
etc.	

We shall now discuss the application of these ideas to a *design procedure.* Suppose we are given the nth-order plant (5.4-1), and a list of $n' < n$ desired dominant poles $z_1, \ldots, z_{n'}$. We shall describe the setting up of an optimal regulator problem such that the associated optimal closed-loop system has dominant poles approximating the desired ones.

We first form the polynomial $m(s) = \prod_{i=1}^{n'} (s - z_i)$. As before, $\psi(s) = \det (sI - F)$. Then an n vector d may be defined by

$$d'(sI - F)^{-1}g = \frac{m(s)}{\psi(s)}, \tag{5.4-12}$$

and using d, we define the matrix Q of the performance index (5.4-9) as $Q = dd'$. We assume $[F, d]$ is completely observable. (See problem 5.4-10 for the contrary case.)

Retaining the definition (5.4-5) of the polynomial $q(s)$, we can check simply that $q(s)$ is precisely $m(-s)m(s)$:

$$\begin{aligned}\frac{q(s)}{\psi(-s)\psi(s)} &= g'(-sI - F')^{-1}Q(sI - F)^{-1}g \\ &= g'(-sI - F')^{-1}dd'(sI - F)^{-1}g \\ &= \frac{m(-s)}{\psi(-s)}\frac{m(s)}{\psi(s)}.\end{aligned}$$

We recall also that for suitably large ρ in the performance index (5.4-9), n' of the optimal closed-loop system poles are approximated by the negative real part zeros of $q(s)$. But since $q(s)$ is $m(-s)m(s)$ and $m(s)$ has negative real part zeros, it follows that for suitably large ρ the optimal closed-loop dominant poles are approximated by the zeros $m(s)$, which are precisely the desired dominant closed-loop poles.

Having fixed the matrix Q of the performance index (5.4-9) as $Q = dd'$, we now ask what the value of ρ should be. The choice of ρ is governed by the need to ensure that the n' zeros of $m(s)$ are truly dominant poles of the closed-loop system—i.e., that the remaining poles of the closed-loop system should have magnitude at least 10 times the magnitude of any of the dominant poles. Now the magnitude of the nondominant poles is approximately $\rho^{1/2(n-n')}$ (as the constant α of our earlier remarks is now unity). Therefore, ρ might be chosen sufficiently large to guarantee that

$$\rho^{1/2(n-n')} \geq 10\,|z_i| \quad i = 1, \cdots, n'. \tag{5.4-13}$$

In summary, the design procedure is as follows:

1. With n' dominant poles, form a polynomial $m(s)$ of degree n', with highest order term $s^{n'}$, and such that $m(s)$ has the desired dominant poles as its zeros.
2. Define the vector d by Eq. (5.4-12).
3. Choose $Q = dd'$ and ρ so that Eq. (5.4-13) is satisfied. The performance index (5.4-9) will be minimized by an optimal control that determines a closed-loop system possessing the approximately desired dominant poles.
4. Determine the feedback law either by the Riccati equation approach or by the procedure given earlier in this section based on factorization of (5.4-10).

To illustrate the procedure, we shall consider *two examples*. First, suppose

$$F = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad g = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and we wish to have a dominant pole of the closed-loop system at $-1 + j0$. (Note that we cannot specify two poles of the closed-loop system.) The polynomial $m(s) = s + 1$ has the single root $s = -1$; now form $m(s)/\psi(s) = (s + 1)/(s^2 + 1)$, and observe that $d' = [1 \quad 1]$ ensures that $d'(sI - F)^{-1}g = m(s)/\psi(s)$. We have also that $n - n' = 1$, and Eq. (5.4-13) yields that $\rho \geq 100$. We shall therefore take $\rho = 100$, and minimize the performance index $\int_{t_0}^{\infty} [u^2 + 100(x_1 + x_2)^2]\, dt$. The closed-loop system poles are the negative real part zeros of

$$\begin{aligned} \psi(-s)\psi(s) + \rho m(-s)m(s) &= s^4 + 2s^2 + 1 - 100s^2 + 100 \\ &= s^4 - 98s^2 + 100 \\ &= s^4 + 20s^2 + 100 - 118s^2 \\ &= (s^2 + \sqrt{118}s + 10)(s^2 - \sqrt{118}s + 10). \end{aligned}$$

Accordingly, the closed-loop poles are the zeros of $s^2 + \sqrt{118}s + 10$, which are computed to be -1.01 and -9.84. The dominant pole is then correct to 1%. The feedback vector k (generating the optimal control according to $u = k'x$) is given by

$$1 - k'(sI - F)^{-1}g = \frac{s^2 + \sqrt{118}s + 10}{s^2 + 1}$$

from which we derive

$$k' = [-9 \quad -\sqrt{118}].$$

For a *second example*, we consider the sixth-order system defined by

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & +1 & +1 & -1 \end{bmatrix} \quad g = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

We aim to have dominant poles at $s = -1$, $s = -1 \pm j1$. The open-loop poles are given by the zeros of $\det (sI - F) = \psi(s)$, which may be computed to be $s^3(s + 1)^2(s - 1) = s^6 + s^5 - s^4 - s^3$. In terms of the earlier notation, the polynomial $m(s)$ is $(s + 1)(s^2 + 2s + 2) = s^3 + 3s^2 + 4s + 2$. By forming $m(s)/\psi(s) = (s^3 + 3s^2 + 4s + 2)/(s^6 + s^5 - s^4 - s^3)$, we can check that $d' = [2 \quad 4 \quad 3 \quad 1 \quad 0 \quad 0]$ ensures that $d'(sI - F)^{-1}g = m(s)/\psi(s)$. We have also that $n - n' = 3$, and so Eq. (5.4-13) yields that $\rho \geq 2^6 \times 10^6$.

However, it turns out that smaller values of ρ will give reasonably acceptable results. Thus, with $\rho = 10^5$, the closed-loop poles are the negative real part zeros of $s^{12} - 3s^{10} + 3s^8 - s^6 + 10^5(-s^6 + s^4 - 4s^2 + 4)$; a computer evaluation of the zeros of this polynomial leads to the closed-loop poles -1.00, $-1.00 \pm j1.00$, -4.72, $-2.35 \pm j3.96$. The dominant poles are therefore correct, at least to within 0.5%, but the remaining poles are not far removed from the dominant poles. The polynomial having zeros that are these closed-loop poles is

$$p(s) = s^6 + 12.4s^5 + 75.5s^4 + 267.5s^3 + 489.5s^2 + 484.4s + 198.5.$$

The feedback vector k is given by $1 - k'(sI - F)^{-1}g = p(s)/\psi(s)$, and is readily found to be $[-198.5\ -484.4\ -489.5\ -268.5\ -76.5\ -11.4]$.

Two warnings are perhaps appropriate at this stage. It will probably not have escaped the reader's notice that the entries of the vector k in the two examples are large, relative to, say, the entries of F and g. This means that the magnitude of the control arising from feeding back $k'x$ may be large, even though x and Fx are small. Equivalently, there will be a tendency for saturation at the plant input to occur as a result of the entries of k being large, and, of course, the larger these entries are, the more likely is the occurrence of saturation. It is not surprising that this phenomenon should occur, in view of the fact that if ρ is large in the performance index (5.4-9), nonzero states are weighted more heavily than nonzero controls. Therefore, larger and larger ρ will tend to make the entries of k larger and larger, and make saturation more and more possible (Problem 5.4-5 asks the student to connect variation in ρ with variation in the entries of k.)

Reference to (5.4-13), the equation that yields the minimum ρ guaranteeing a dominancy factor of 10, shows that the more nondominant poles there are, the larger ρ must be. Therefore, the possibility of saturation is lowered by having, say, only one nondominant closed-loop pole.

The second warning concerns the interpretation of dominancy. Suppose a transfer function is given with n poles, n' of which are dominant. Suppose also that there are n'' dominant zeros, and suppose we are interested in the step response of a system with this transfer function, Then, unless $n'' \leq n'$, or, better, $n'' < n'$, the residues associated with the nondominant poles will not be small relative to the residues associated with the dominant poles. Accordingly, the step response will only be free of a sizable component of fast transients if the upper bound on n'' is met. Hence, the term nondominant pole really has the fuller meaning: the associated component of transient response is not only very fast, and therefore dies away quickly, but is also small, and therefore negligible. We conclude that any attempt at pole positioning to obtain satisfactory transient response should take into account the relative number of dominant zeros and dominant poles. [The effect of state-variable feedback on the zeros of a transfer function will be discussed shortly.]

We shall now discuss a straightforward *extension of the closed-loop pole specification idea to multiple-input systems.* To apply the preceding design procedure, we adopt an ad hoc procedure for converting the multiple-input system into a single-input system.

The procedure relies on an observation of [5] that if $[F, G]$ is completely controllable, where F is $n \times n$ and G is $n \times m$, and if F has distinct eigenvalues, then $[F, G\gamma]$ is "almost surely" completely controllable, where γ is an arbitrary nonzero m vector. If F does not have distinct eigenvalues, the observation *may* still be true. There is, of course, a technical meaning for "almost surely," which we shall not bother about here. We shall take the term to mean what we would normally understand by it. That is, with $[F, G]$ completely controllable, if we choose an arbitrary γ such that by some rare chance $[F, G\gamma]$ is not completely controllable, then an infinitesimal change in γ will ensure that $[F, G\gamma]$ is completely controllable.

Returning to the closed-loop pole specification problem, let us suppose we are given a completely controllable m input system that

$$\dot{x} = Fx + G\bar{u}, \tag{5.4-14}$$

where we label the input $\bar{u}$ to distinguish it from the input u of a single-input system, which we define in a moment. Let γ be an arbitrary m vector, such that $[F, G\gamma]$ is completely controllable. The equation

$$\dot{x} = Fx + G\gamma u \tag{5.4-15}$$

defines a completely controllable single-input system. For the moment, we focus attention on (5.4-15) rather than (5.4-14). We can find a large constant ρ, and an n vector d (which then defines a matrix $Q = dd'$), such that the optimal control law $u = k'x$ minimizing the performance index (5.4-9) gives an optimal closed-loop system

$$x = (F + G\gamma k')x \tag{5.4-16}$$

with closed-loop poles approximating the closed-loop poles desired for the m-input system.

Now we observe that if the control law $\bar{u} = \gamma k'x$ is used for the open-loop scheme (5.4-14), the resulting closed-loop scheme is also (5.4-16), and, consequently, the desired closed-loop poles are achieved. The control law $\bar{u} = \gamma k'x$ is not strictly an optimal control for (5.4-14), but in the light of the closed-loop scheme (5.4-16) being optimal [admittedly, for a different open-loop system than (5.4-14), but this is somewhat irrelevant], it would be expected that some of the ancillary benefits of optimality, including those yet to be discussed, would accrue.

Problem 5.4-7 asks for a design scheme for generating desired closed-loop poles, which also results in a truly optimal multiple-input closed-loop system.

A second, but sometimes unsatisfactory, *approach to achieving desired*

closed-loop pole-positions is the following. We suppose we are given a completely controllable system with no restriction on the number of inputs:

$$\dot{x} = Fx + Gu. \tag{5.4-17}$$

Furthermore, given a set of desired closed-loop poles, the number of poles being equal to the dimension of F, we construct a matrix $\bar{F}$ the set of eigenvalues of which coincides with the set of the these poles. We then form the following performance index:

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} [u'Ru + \rho(\dot{x} - \bar{F}x)'Q(\dot{x} - \bar{F}x)]\,dt. \tag{5.4-18}$$

Here, R, Q, and ρ satisfy the usual constraints. The presence of the second term in the integrand of the performance index (5.4-18) serves to force the optimum state trajectory to be close to the solutions of $\dot{x} = \bar{F}x$, which amounts to forcing the closed-loop poles to be near to the eigenvalues of $\bar{F}$.

To minimize (5.4-18) is quite straightforward. First, the quantity $\dot{x}$ in the integrand is replaced by $Fx + Gu$, using (5.4-17), which allows the integrand to be written in the form

$$u'R_1u + 2x'S_1u + x'Q_1x,$$

with $Q_1 - S_1R_1^{-1}S_1'$ nonnegative definite and R_1 positive definite. With this rewriting of the integrand, determination of the control law is straightforward, using the ideas of Chapter 3, Sec. 3.4.

The minimizing of (5.4-18) will lead to closed-loop poles that approximate, but in general do not equal, the desired values. As ρ becomes larger, the degree of approximation will clearly improve.

The potential disadvantage of this form of optimal control lies in the presence of the cross product term $x'S_1u$ in the integrand of the performance index. As noted in the last problem of the immediately preceding section, the presence of such a term may destroy the pleasing properties associated with phase margin and gain margin. It also destroys nearly all the other pleasing engineering properties to be discussed in the next two chapters.

In the remainder of this section, we shall consider *what happens to the zeros of a system where optimal linear feedback of the states is introduced.* We consider the situation of Fig. 5.1-2, where the prescribed plant has transfer function $h'(sI - F)^{-1}g$, and a feedback $k'x$ is applied. The transfer function between input and output after applying feedback is $h'(sI - F - gk')^{-1}g$. Therefore, we are concerned with the answer to the following question: How do the zeros of the transfer function $h'(sI - F - gk')^{-1}g$ differ from those of the transfer function $h'(sI - F)^{-1}g$? The answer is that the zeros of the two transfer functions are the same, as we shall now show.

Suppose that $\det(sI - F) = \psi(s)$ and that $\det(sI - F - gk') = p(s)$, and let polynomials $l_1(s)$ and $l_2(s)$ be defined by

$$h'(sI - F)^{-1}g = \frac{l_1(s)}{\psi(s)} \qquad h'(sI - F - gk')^{-1}g = \frac{l_2(s)}{p(s)}. \tag{5.4-19}$$

We have to show that $l_1(s) = l_2(s)$. Now

$$\begin{aligned} h'(sI - F - gk')^{-1}g &= h'(sI - F - gk')^{-1}(sI - F)(sI - F)^{-1}g \\ &= h'(sI - F - gk')^{-1}[(sI - F - gk') + gk'](sI - F)^{-1}g \\ &= h'(sI - F)^{-1}g + h'(sI - F - gk')^{-1}gk'(sI - F)^{-1}g. \end{aligned}$$

Therefore,

$$h'(sI - F - gk')^{-1}g = \frac{h'(sI - F)^{-1}g}{1 - k'(sI - F)^{-1}g}. \tag{5.4-20}$$

Now the definition we have first adopted for $p(s)$ coincides with that of (5.4-6), and with $p(s)$ so defined, Eq. (5.4-4), repeated here for convenience, is also valid:

$$1 - k'(sI - F)^{-1}g = \frac{p(s)}{\psi(s)}. \tag{5.4-21}$$

When Eqs. (5.4-19) and (5.4-21) are inserted into (5.4-20), the result $l_1(s) = l_2(s)$ is immediate.

In reality, the optimality of k is not a relevant factor in the preceding argument. Thus, we have established the important result that state feedback (optimal or nonoptimal) leaves invariant the zeros of the open-loop transfer function.

As a note of warning, we draw attention to the need not to confuse the transfer function $h'(sI - F)^{-1}g$ of Fig. 5.1-2 with the transfer function of the open loop part of the overall scheme, viz., $-k'(sI - F)^{-1}g$, some of the properties of which we studied at some length in the last section. Although the zeros of this transfer function and its closed-loop counterpart $-k'(sI - F - gk)^{-1}g$ are the same, they are, of course, both dependent on k.

As references for the material of this section, we may list [1] and [6], which discuss the properties of the closed-loop system poles when ρ approaches infinity in the performance index (5.4-9); [7], which discusses the second pole-positioning procedure presented; and [8], which demonstrates the invariance of the zeros of a transfer function with state-variable feedback. Chapter 7, Sec. 7.4, considers further the problem of pole positioning in optimal systems.

Problem 5.4-1. Consider a system

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u$$

with performance index $\int_{t_0}^{\infty} (u^2 + x_1^2 + x_2^2)\,dt$. Using the polynomial factorization technique, find the optimal control.

Problem 5.4-2. With the same system as in Problem 5.4-1, suppose the performance index is now $\int_{t_0}^{\infty} (u^2 + 900x_1^2 + 900x_2^2)\,dt$. What are approximate optimal closed-loop poles? Compare the approximate pole positions with the actual ones.

Problem 5.4-3. Consider a system with transfer function $w(s) = 1/s^3(s+1)^3$. Write down a 6×6 matrix F, and vectors g and h such that $w(s) = h'(sI - F)^{-1}g$. Then define and solve an optimization problem so that with state variable feedback, the new system transfer function becomes approximately $a/(s+1)(s^2+s+1)$ for some constant a. [*Hint:* A choice for F of the companion matrix form and for g' of the form $[0 \cdots 0\ 1]$ may be straightforward.]

Problem 5.4-4. Consider a system $\dot{x} = Fx + gu$ with performance index $\int_{t_0}^{\infty} (u^2 + \rho x'Qx)\, dt$ for nonnegative definite Q and positive ρ. Suppose that with D any matrix such that $DD' = Q$, the pair $[F, D]$ is completely observable. Discuss what happens to the poles of the closed-loop system as ρ approaches zero. Let $u = k'x$ be an optimal control law resulting for very small ρ. Discuss the difference between the magnitude and phase of the transfer functions $h'(j\omega I - F)^{-1}g$ and $h'(j\omega I - F - gk')^{-1}g$ as ω varies.

Problem 5.4-5. Suppose $u = k'x$ is the optimal control for $\dot{x} = Fx + gu$, with performance index $\int_{t_0}^{\infty} (u^2 + \rho x'Qx)\, dt$. Show that as ρ approaches infinity, some, if not all, the entries of k approach infinity as fast as $\sqrt{\rho}$.

Problem 5.4-6. Consider the multiple-output, multiple-input system $\dot{x} = Fx + Gu$, $y = H'x$. Suppose state-variable feedback is applied, so that the new closed-loop system equations are $\dot{x} = (F + GK')x + Gu$, $y = H'x$. Can you indicate quantities analogous to the zeros of a scalar transfer function which remain invariant with the specified feedback?

Problem 5.4-7. Consider the multiple-input system $\dot{x} = Fx + Gu$, with performance index $\int_{t_0}^{\infty} (u'u + \rho x'Qx)\, dt$, with the usual constraints on Q. If $u = K'x$ is the optimal control law, it follows that $[I - G'(-sI - F')^{-1}K][I - K'(sI - F)^{-1}G] = I + \rho G'(-sI - F')^{-1}Q(sI - F)^{-1}G$. Now, because of the easily established relation

$$[I - K'(sI - F)^{-1}G]^{-1} = I + K'(sI - F - GK')^{-1}G,$$

it is evident that the poles of the closed-loop system are the zeros of $\det [I - K'(sI - F)^{-1}G]$. Assuming that, except for isolated values of s, $G'(-sI - F')^{-1}Q(sI - F)^{-1}G$ is nonsingular, discuss a technique for choosing Q and ρ so that the closed-loop system has certain desired poles. (This is a difficult problem. See [9] for a lengthy discussion.)

Problem 5.4-8. In this problem, a relation is developed between the bandwidth of a single-input, single-output, closed-loop, optimal system and some of the parameters appearing in the optimization problem. This allows design for fixed bandwidth of an optimal system, [10]. Consider the transfer function $w(s) = h'(sI - F)^{-1}g$, with $[F, g]$ completely controllable and $[F, h]$ completely observable. Suppose a state feedback law is found that minimizes the performance index,

$$\int_{t_0}^{\infty} [u^2 + \rho(h'x)^2]\, dt.$$

Let $w_{\text{CL}}(s)$ denote the corresponding closed-loop transfer function, and let ω_0 be

the frequency for which

$$|w_{CL}(j\omega_0)|^2 = \tfrac{1}{2}|w_{CL}(0)|^2.$$

Show that ω_0 may be determined as the frequency for which

$$|w(j\omega_0)|^2 = \frac{1}{\rho + (2/|w(0)|^2)}.$$

[Note the simple form of this result for the case when $w(s)$ has a pole at $s = 0$.] Explain what happens as ρ approaches infinity, considering first the case when $w(s)$ is minimum phase—i.e., its poles and zeros all have negative real parts.

Problem 5.4-9. Suppose you are given a single-input, single-output system,

$$\dot{x} = Fx + gu \qquad y = h'x.$$

Suppose also that it is desired to have the output behave in accordance with the law $\dot{y} = -\alpha y$ for a certain positive α. Consider the effect of using the performance index

$$V = \int_{t_0}^{\infty} [u^2 + \rho(\dot{y} + \alpha y)^2]\, dt.$$

Explain what happens as $\rho \to \infty$, considering the following two cases.

1. The zeros of $h'(sI - F)^{-1}g$ are in $Re[s] < 0$.
2. The zeros of $h'(sI - F)^{-1}g$ are arbitrary.

Problem 5.4-10. Discuss the case when the polynomial $m(s)$ in Eq. (5.4-12) is such that $[F, d]$ is not completely observable. [*Hint:* Use the fact that the dynamics of the unobservable states are defined by common zeros of $m(s)$ and $\psi(s)$.]

REFERENCES

[1] Kalman, R. E., "When is a Linear Control System Optimal?" *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 86, March 1964, pp. 1–10.

[2] Kreindler, E., "Closed-Loop Sensitivity Reduction of Linear Optimal Control Systems," *IEEE Trans. Auto. Control*, Vol. AC–13, No. 3, June 1968, pp. 254–262.

[3] Anderson, B. D. O., "The Inverse Problem of Optimal Control," *Rep. No. SEL–66–038* (*TR No. 6560–3*), May 1966, Stanford Electronics Laboratories, Stanford, Calif.

[4] Thaler, G. J., and R. G. Brown, *Analysis and Design of Feedback Control Systems*, 2nd ed., McGraw-Hill Book Company, New York, 1960.

[5] Gopinath, B., and B. O. Lange, "On the Identification and Control of Linear Systems," *Tech. Rep. SUDAAR No. 351*, June 1968, Dept. of Aeronautics and Astronautics, Stanford Univ., Calif.

[6] Chang, S. S. L. *Synthesis of Optimum Control Systems*, McGraw-Hill Book Company, New York, 1961.

[7] Kalman, R. E., and T. S. Englar, "A User's Manual for ASPC," *RIAS Rep. NASA*, 1965, Baltimore, Md.

[8] Brockett, R. W., "Poles, Zeros and Feedback: State Space Interpretation," *IEEE Trans. Auto. Control*, Vol. AC-10, No. 2 April 1965, pp. 129–135.

[9] Tyler, J. S., and F. B. Tuteur, "The Use of a Quadratic Performance Index to Design Multivariable Control Systems," *IEEE Trans. Auto. Control*, Vol. AC-11, No. 1, January 1966, pp. 84–92.

[10] Leake, R. J., "Return Difference Bode Diagrams for Optimal System Design," *IEEE Trans. Auto. Control*, Vol. AC-10, No. 3, July 1965, pp. 342–344.

CHAPTER 6

INSERTION OF NONLINEARITIES IN OPTIMAL REGULATORS

6.1 TOLERANCE OF NONLINEARITIES IN THE FEEDBACK LOOP

Many nonlinear physical systems may be described by linear models; analysis may then be carried out on the linear model, and the conclusions of this analysis may be used as an approximate guide to the behavior of the actual nonlinear system. If the nonlinearity of the physical system is small in some sense, then it would be expected the physical system and the model would behave similarly. If the nonlinearity of the physical system is not small, it is, of course, still desirable and sometimes possible to make predictions, based on analysis of an associated linear model, which will give information about the physical system.

In this section, we shall consider the introduction of gross nonlinearities into regulator systems, and we shall be concerned with establishing stability properties of the resulting nonlinear systems from certain properties that we know about (linear) regulator systems.

The scheme we consider is shown in Fig. 6.1-1. We suppose we are given a completely controllable system,

$$\dot{x} = Fx + Gu. \tag{6.1-1}$$

We suppose also that there is an associated performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} e^{2\alpha t}(u'u + x'Qx)\, dt \tag{6.1-2}$$

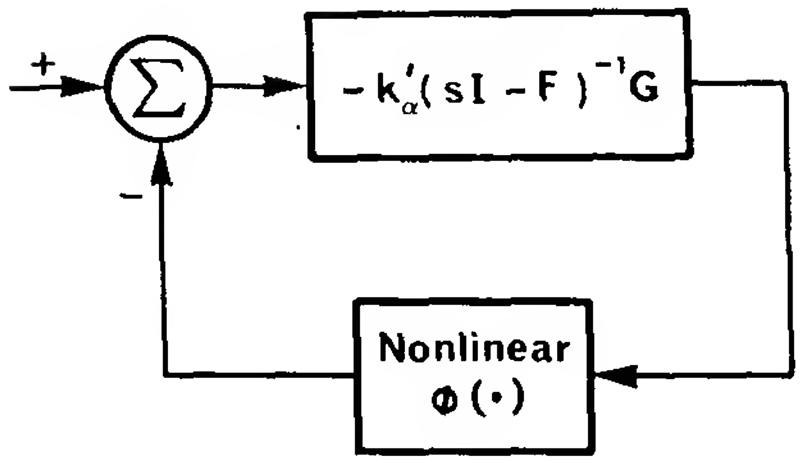


Fig. 6.1-1 Optimal system with inclusion of nonlinearities.

where α is a nonnegative constant, Q is nonnegative definite symmetric, and if D is any matrix such that $Q = DD'$, the pair $[F, D]$ is completely observable. [We could, of course, consider a more general performance index than (6.1-2)—one where $u'u$ is replaced by $u'Ru$. But we can note that the associated minimization problem is equivalent to one of the form defined by (6.1-1) and (6.1-2), because the minimization of $\int_{t_0}^{\infty} e^{2\alpha t}(u'Ru + x'Qx)\,dt$ subject to $\dot{x} = Fx + Gu$ is equivalent to the minimization of $\int_{t_0}^{\infty} e^{2\alpha t}(u_1'u_1 + x'Qx)\,dt$ subject to $\dot{x} = Fx + GR^{-1/2}u_1$ under the identification $u_1 = R^{1/2}u$.] The matrix K_α appearing in Fig. 6.1-1 defines the optimal control for the minimization problem of (6.1-1) and (6.1-2)—i.e.,

$$u = K'_\alpha x \tag{6.1-3}$$

minimizes (6.1-2). We recall that the matrix K_α is given by

$$K_\alpha = -P_\alpha G, \tag{6.1-4}$$

where P_α is the unique positive definite solution of

$$P_\alpha(F + \alpha I) + (F' + \alpha I)P_\alpha - P_\alpha GG'P_\alpha + Q = 0. \tag{6.1-5}$$

Figure 6.1-1, of course, shows a variant on the optimal control arrangement. If the nonlinear feedback block were, in fact, a block that had the same outputs as inputs, Fig. 6.1-1 would represent a truly optimal system. But we allow a variation from the optimal arrangement; we shall permit a wide class of nonlinearities, to be defined quantitatively.

If the external input to the arrangement of Fig. 6.1-1 is zero, it is clear that there would be no difference if the nonlinearities were moved to the right of the summing point—i.e., between the summing point and the block with transfer function matrix $-K'_\alpha(sI - F)^{-1}G$. They could then be regarded as being located at the inputs of the system $\dot{x} = Fx + Gu$ rather than in the feedback loop. Both interpretations are equally valid. Nonlinearities in the input transducers of physical systems are quite common (e.g., because commonly high power levels are present, and saturation may occur).

We shall now define the class of nonlinearities permitted. Then we shall discuss the stability of the system (6.1-1) with nonlinear feedback. The nonlinear block $\phi(\cdot)$ will transform an m vector y, where m is the dimension of

u, into an m vector $\phi(y)$. In addition, we require that $\phi(y)$ be continuous and that the following constraints on $\phi(\cdot)$ hold for all y:

$$(\tfrac{1}{2} + \epsilon)y'y \leq y'\phi(y), \tag{6.1-6}$$

where ϵ is an arbitrary small positive number. This is not an unreasonable condition to investigate, because the gain margin properties of regulators discussed in a previous chapter indicated (at least for single-input systems) that if $\phi(\cdot)$ were linear, then it could satisfy (6.1-6) without a loss in stability of the closed-loop system. Note that (6.1-6) does *not* require the ith output of the nonlinearity $\phi_i(y)$ to depend purely on the ith input y_i; in other words, the nonlinearity can somehow "mix up" its inputs. Nevertheless, the situation where $\phi_i(y)$ depends solely on y_i for each i is the most common. Then (6.1-6) becomes

$$\frac{1}{2} + \epsilon \leq \frac{\phi_i(y_i)}{y_i} \qquad i = 1, 2, \ldots, m \tag{6.1-7}$$

when $y_i \neq 0$. Equation (6.1-7) implies that the graph of $\phi_i(y_i)$ versus y_i for each i must be within any sector bounded by straight lines of slope $\frac{1}{2}$ and ∞; in Fig. 6.1-2, the shaded sector is a permissible region in which the

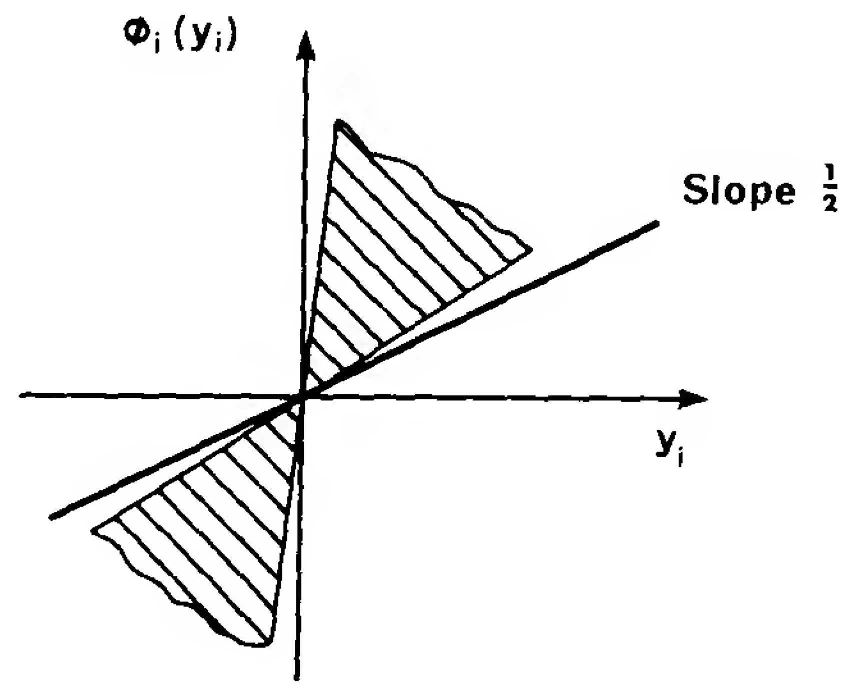


Fig. 6.1-2 Permissible region for graph $\phi_i(y_i)$ versus y_i.

graph of $\phi_i(y_i)$ versus y_i should lie. Notice that this figure illustrates that $\phi_i(0) = 0$ for each i—in fact, the more general constraint (6.1-6) and the continuity requirement imply that $\phi(0) = 0$. Now, with no external input, the scheme of (6.1-1) is represented by

$$\dot{x} = Fx + G\phi(K'_\alpha x). \tag{6.1-8}$$

We need to examine the stability of (6.1-8) for all initial conditions. Let us take as a Lyapunov function

$$V(x) = x'P_\alpha x \tag{6.1-9}$$

and observe that $V(x)$ satisfies the requirements listed in Theorems A and

B of Chapter 3, Sec. 3.2, viz., $V(x)$ is positive definite and approaches infinity as $x'x$ approaches infinity.

Next, we need to evaluate $\dot{V}$. We have (with the various steps to be explained),

$$\begin{aligned}
\dot{V}(x) &= \dot{x}'P_\alpha x + x'P_\alpha \dot{x} \\
&= x'(P_\alpha F + F'P_\alpha)x + 2x'P_\alpha G\phi(K'_\alpha x) \\
&= x'[P_\alpha(F + \alpha I) + (F' + \alpha I)P_\alpha - P_\alpha GG'P_\alpha + Q]x \\
&\quad -2\alpha x'P_\alpha x - x'Qx + x'P_\alpha GG'P_\alpha x + 2x'P_\alpha G\phi(-G'P_\alpha x) \\
&= -2\alpha x'P_\alpha x - x'Qx + x'P_\alpha GG'P_\alpha x + 2x'P_\alpha G\phi(-G'P_\alpha x) \\
&\leq -2\alpha x'P_\alpha x - x'Qx + x'P_\alpha GG'P_\alpha x - 2(\tfrac{1}{2} + \epsilon)x'P_\alpha GG'P_\alpha x \\
&= -2\alpha x'P_\alpha x - x'Qx - 2\epsilon x'P_\alpha GG'P_\alpha x.
\end{aligned} \tag{6.1-10}$$

The second equality follows from (6.1-8), the third by simple rearrangement and the relation $P_\alpha G = -K_\alpha$, and the fourth from Eq. (6.1-5). The inequality follows from (6.1-6), and the final equality again by rearrangement. Consequently,

$$\dot{V}(x) \leq -x'Qx - 2\alpha x'P_\alpha x - 2\epsilon x'K_\alpha K'_\alpha x \tag{6.1-11}$$

At this stage, we shall separate our discussion of the cases $\alpha = 0$ and $\alpha > 0$. First, for $\alpha = 0$, we have from (6.1-11) that

$$\dot{V}(x) \leq -x'Qx - 2\epsilon x'K_0K'_0x. \tag{6.1-12}$$

Certainly, therefore, $\dot{V}$ is nonpositive, and thus (6.1-8) is at least stable. But we shall also argue that $\dot{V}(x)$ is not identically zero along a nonzero trajectory of (6.1-8), and thus asymptotic stability follows by Theorem B, Chapter 3, sec. 3.2. The argument is by contradiction. If $\dot{V}(x) \equiv 0$ from some time t_1 on, then $K'_0x \equiv 0$ from t_1 on, and thus $\phi(K'_0x) \equiv 0$. Then (6.1-8) becomes the same as $\dot{x} = Fx$. But $\dot{V}(x) \equiv 0$ also implies $x'Qx \equiv 0$, and thus $x'(t_1)$ $\exp[F'(t - t_1)]Q \exp[F(t - t_1)]x(t_1) = 0$ for all $t \geq t_1$. This contradicts the complete observability of the pair $[F, D]$, where D satisfies $DD' = Q$. Therefore, $\dot{V}(x) \equiv 0$ for $t \geq t_1$ is impossible, and asymptotic stability truly prevails.

Now consider the case $\alpha > 0$. From (6.1-11) we have immediately

$$\dot{V}(x) \leq -2\alpha x'P_\alpha x = -2\alpha V(x), \tag{6.1-13}$$

which implies that $V(x(t)) \leq e^{-2\alpha t}V(x(0))$ by the following simple argument. Let $a(t)$ be a nonnegative function such that $\dot{V}(x) = -2\alpha V(x) - a(t)$. The solution of this differential equation is well known to be $V(x(t)) = e^{-2\alpha t}V(x(0)) - \int_0^t e^{-2\alpha t}e^{2\alpha t}a(\tau)\,d\tau$, from which the desired conclusion is immediate. Rewriting $V(x(t)) \leq e^{-2\alpha t}V(x(0))$ as

$$x'(t)P_\alpha x(t) \leq e^{-2\alpha t}x'(0)P_\alpha x(0), \tag{6.1-14}$$

we see that since $x(\cdot)$ is the solution of (6.1-8), then, using an obvious definition of degree of stability, (6.1-8) possesses degree of stability α.

We recall that with $\alpha = 0$, the optimal system is known to be asymptotically stable, and with $\alpha > 0$, the optimal system is known to have degree of stability α. Thus, the introduction of the nonlinearity $\phi(\cdot)$ does not affect these stability properties of the closed-loop system.

For single-input systems, there is a connection with the material of Chapter 5, Sec. 5.3, where we discussed the question of the gain margin of the optimal regulator. We recall from that section that if a constant (linear) gain $\beta > \frac{1}{2}$ is placed in the feedback loop of an optimal regulator, the stability properties of the regulator are not disturbed. Now a constant linear gain greater than $\frac{1}{2}$ is a special case of a *nonlinearity* confined to the sector of Fig. 6.1-2. Therefore, the earlier result of Chapter 5 is included within the more general framework of this section.

We shall now show how the constraint of (6.1-6) may sometimes be relaxed even further, in the sense that stability results may follow for the following revised constraint:

$$\beta y'y \leq y'\phi(y) \tag{6.1-15}$$

for certain $\beta \leq \frac{1}{2}$. Since the constraint $\beta > \frac{1}{2}$ for the case when α is nonzero leads to a closed-loop system with degree of stability α, we would expect that for a certain range of $\beta \leq \frac{1}{2}$, asymptotic stability would still prevail, but perhaps with a smaller degree of stability. Indeed, we shall verify this property.

For the moment, we do not restrict α. As before, we consider the closed-loop system (6.1-8), with Lyapunov function $V(x) = x'P_\alpha x$. The second equality in (6.1-10) follows immediately. We repeat it here for convenience:

$$\dot{V}(x) = x'(P_\alpha F + F'P_\alpha)x + 2x'P_\alpha G\phi(K'_\alpha x). \tag{6.1-16}$$

Using $P_\alpha G = -K_\alpha$, and the new constraint (6.1-15), we have

$$\dot{V}(x) \leq x'(P_\alpha F + F'P_\alpha)x - 2\beta x'K_\alpha K'_\alpha x$$

from which we can derive, using (6.1-5),

$$\dot{V}(x) \leq -x'Qx - 2\alpha x'P_\alpha x + (1 - 2\beta)x'K_\alpha K'_\alpha x. \tag{6.1-17}$$

Now we shall consider separately the cases $\alpha = 0$, $\alpha > 0$. When $\alpha = 0$, (6.1-17) is

$$\dot{V}(x) \leq -x'[Q - (1 - 2\beta)K_0 K'_0]x. \tag{6.1-18}$$

Now, if $\beta \leq \frac{1}{2}$ is chosen so that $Q - (1 - 2\beta)K_0 K'_0$ is nonnegative definite, stability of (6.1-8) will follow, and if $Q - (1 - 2\beta)K_0 K'_0$ is positive definite, asymptotic stability will follow. Do such β exist? Possibly not, but if Q is positive definite, certainly there are values of β *less* than $\frac{1}{2}$ for which $Q - (1 - 2\beta)K_0 K'_0$ is positive definite. For a symmetric matrix A, let $\lambda_{\max}(A)$

denote the maximum eigenvalue and $\lambda_{\min}(A)$ the minimum eigenvalue; clearly, $\lambda_{\max}(A)I \geq A \geq \lambda_{\min}(A)I$. Then a sufficient condition on β ensuring that $Q - (1 - 2\beta)K_0K_0'$ is positive definite is

$$\lambda_{\min}(Q) - (1 - 2\beta)\lambda_{\max}(K_0K_0') > 0. \tag{6.1-19}$$

But if, on the other hand, Q is singular, then there may be no β less than $\frac{1}{2}$ for which $Q - (1 - 2\beta)K_\alpha K_\alpha'$ is even nonnegative definite, although with $\beta = \frac{1}{2}$, nonnegative definiteness follows.

Now consider the case $\alpha > 0$. Equation (6.1-17) becomes

$$\dot{V}(x) \leq -x'[Q + 2\alpha P_\alpha - (1 - 2\beta)K_\alpha K_\alpha']x \tag{6.1-20}$$

and certainly now (because P_α is positive definite) there exist β less than $\frac{1}{2}$ such that $Q + 2\alpha P_\alpha - (1 - 2\beta)K_\alpha K_\alpha'$ is positive definite. A sufficient condition on β that ensures this is that

$$\lambda_{\min}(Q + 2\alpha P_\alpha) - (1 - 2\beta)\lambda_{\max}(K_\alpha K_\alpha') > 0. \tag{6.1-21}$$

If for some fixed β, $Q + 2\alpha P_\alpha - (1 - 2\beta)K_\alpha K_\alpha'$ is positive definite, it is possible to compute a lower bound on the resultant degree of stability of the closed-loop system. Let γ be such that

$$\lambda_{\min}[Q + 2\alpha P_\alpha - (1 - 2\beta)K_\alpha K_\alpha'] = \gamma\lambda_{\max}(P_\alpha). \tag{6.1-22}$$

Then (6.1-20) implies $\dot{V}(x) \leq -\gamma\lambda_{\max}(P_\alpha)x'x \leq -\gamma x'P_\alpha x$—i.e., $\dot{V}(x) \leq -\gamma V(x)$. So $\gamma/2$ is a lower bound on the degree of stability of the closed-loop system.

Smaller β lead to smaller γ; thus, as β becomes smaller, there is a continuous exchange between the amount of tolerable nonlinearity and the degree of stability. This exchange eventually terminates when the amount of tolerable nonlinearity exceeds that which will permit stability to be retained.

Implicitly in the foregoing discussion, we have assumed that the nonlinearity ϕ is time invariant—i.e., for all t_1 and t_2, $\phi(y(t_1)) = \phi(y(t_2))$ if $y(t_1) = y(t_2)$. But a speedy review of the calculations will soon show that time invariance of $\phi(\cdot)$ is not essential; it may just as well be time varying. Thus, in the special case of a single-input system, so long as a time-varying nonlinear gain $\phi(y, t)$ has the property that $\frac{1}{2} + \epsilon \leq \phi(y, t)/y$ for all nonzero y, stability of the closed-loop system follows.

It is possible to view this, and indeed the corresponding multiple-input result, as following from an application of the *circle criterion* [1]. To state this criterion, consider the feedback arrangement of Fig. 6.1-3(a) and assume the external input is zero. Suppose also that the time-varying nonlinear gain element with input y has output $\kappa(t, y)y$, and that the graph of $\kappa(t, y)y$ versus y for all t is confined to a sector itself inside a sector defined by straight lines of slope κ_1 and κ_2 with $0 \leq \kappa_1 < \kappa_2$, as shown in Fig. 6.1-3(b). The circle criterion states that if the closed-loop system is asymptotically stable for all

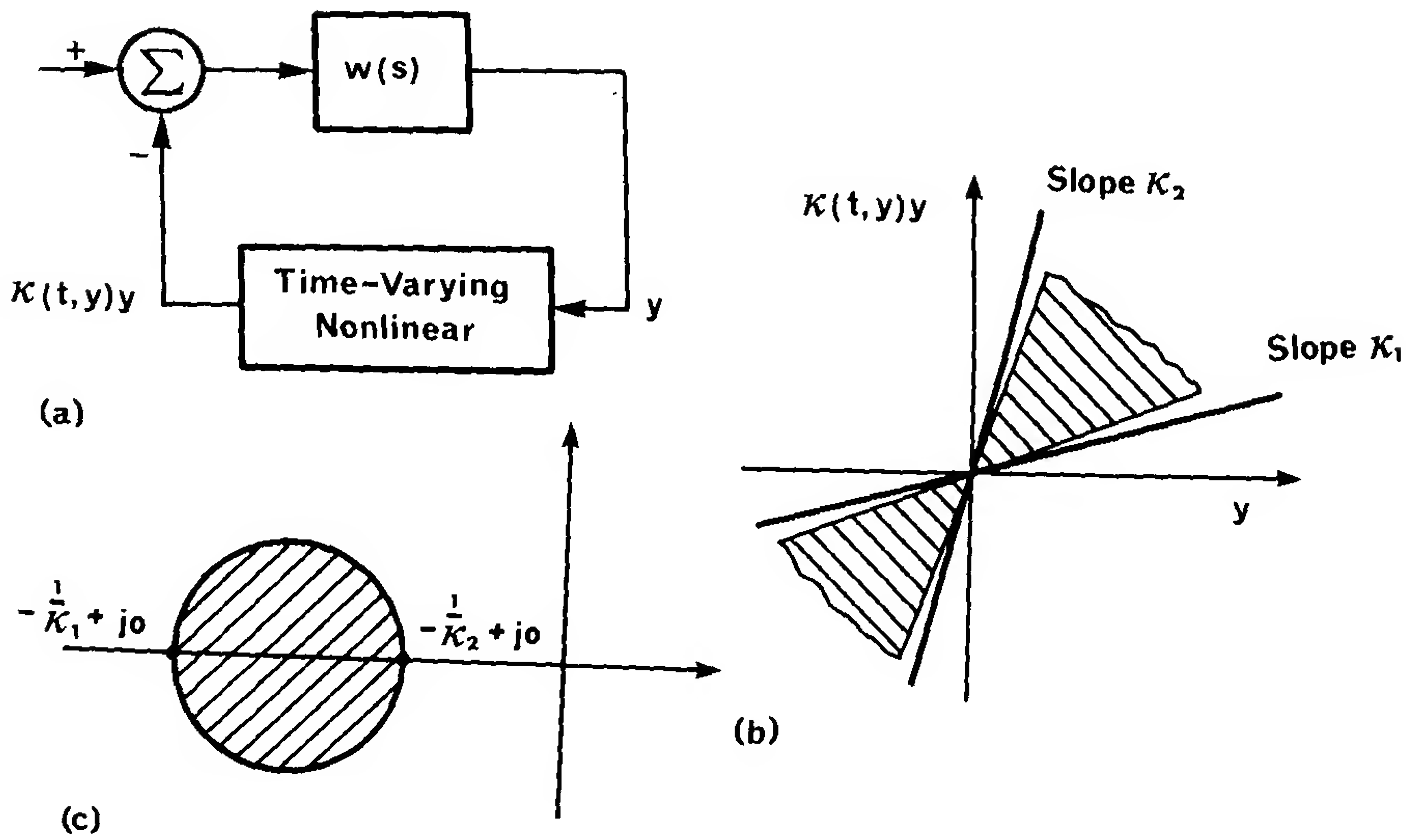


Fig. 6.1-3 Diagrams illustrating application of circle criterion.

constant linear gains κ with $\kappa_1 < \kappa < \kappa_2$, and if the Nyquist plot of $w(j\omega)$ does not enter the circle on a diameter defined by the points $-(1/\kappa_1) + j0$, $-(1/\kappa_2) + j0$, as shown in Fig. 6.1-3(c), then the closed-loop system with $\kappa(t, y)$ as described is asymptotically stable.

In the special case of an optimal regulator, we recall from Chapter 5, Sec. 5.3, that the Nyquist plot of $-k'_\alpha(j\omega I - F)^{-1}g$ does not enter the circle on diameter $-2 + j0, 0$, and at the same time the closed-loop system of Fig. 6.1-4 with zero external input is asymptotically stable for all constant linear

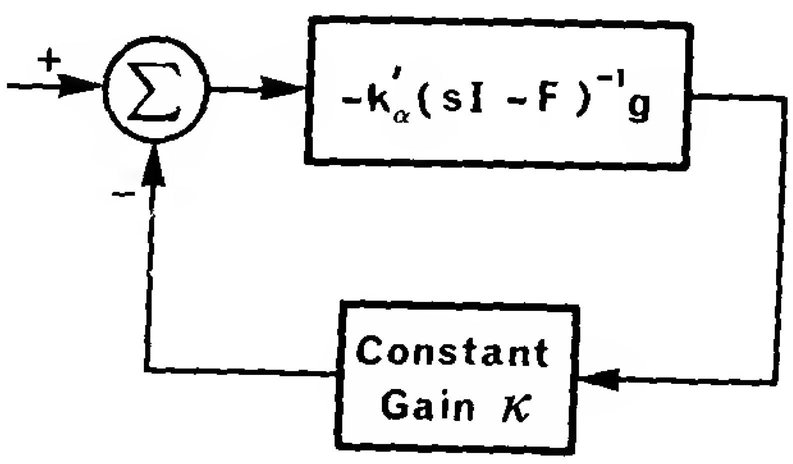


Fig. 6.1-4 Closed-loop system with feedback gain κ.

gains $\kappa > \frac{1}{2}$. Therefore, the requirements of the circle criterion are satisfied with $\kappa_1 = \frac{1}{2}$, $\kappa_2 = \infty$, and the conclusion is that nonlinear time-varying gains can be tolerated of the type described earlier in our Lyapunov function analysis of stability.

There is another important result concerning the exchange of degree of stability and tolerance of nonlinearities $\phi(\cdot)$ or time-varying gain elements $\kappa(t)$. If $\phi(\cdot)$ is considered as a time-varying gain, then the degree of stability

can be exchanged with rate of change of gain variations. From [2] and [3], we shall quote general results that may then be applied directly to the optimal control case.

Using the notation introduced for a description of the circle criterion, we assume that the closed-loop system $w(s)/(1 + \kappa w(s))$ has a degree of stability α for all constant linear gains κ with $0 \leq \kappa_1 < \kappa \leq \kappa_2$ for some constants κ_1 and κ_2. We introduce the quantity $\hat{k}$ defined for fixed but arbitrary T by

$$\hat{k} = \sup_{t \geq 0} \frac{1}{4T} \int_t^{t+T} \left| \frac{\dot{\kappa}(\tau)(\kappa_2 - \kappa_1)}{[\kappa(\tau) - \kappa_2][\kappa(\tau) - \kappa_1]} \right| d\tau.$$

This quantity $\hat{k}$ is in some way a measure of the time variation of $\kappa(t)$. The key result is that for $\hat{k} < \sigma \leq \alpha$, the degree of stability of the closed-loop system is $(\alpha - \sigma)$. This result provides a direct and continuous trade-off between degree of stability and tolerance of time variation. It has immediate application to the optimal regulator with a prescribed degree of stability α, where by taking $\kappa_1 < \frac{1}{2}$ and $\kappa_2 = \frac{1}{2}$, we can perhaps extend the regions of known degree of stability.

So far in this section, the presence of nonlinearities has been considered at one point only of an optimal control system (provided we regard input nonlinearities as being the same as feedback nonlinearities). The natural question arises as to whether any remarks can be made concerning the effect of having nonlinearities at other points of a nominally optimal system. Certainly, a "small" amount of nonlinearity can be tolerated anywhere, without affecting asymptotic stability, when the truly optimal system is itself asymptotically stable. This is a basic result of stability theory (e.g., see [4]). Reference [5] actually computes sector bounds on arbitrary nonlinearities, which, if not exceeded, will retain asymptotic stability.

The proof of the result that gross input nonlinearities are tolerable for optimal regulators will be found in [6] for the scalar input case. The vector input case is covered in [5], which indicates the connection with the circle criterion for multiple-input systems. Results are also possible for time-varying systems, for which [7] should be consulted.

Problem 6.1-1. Consider a single-input, completely controllable system $\dot{x} = Fx + gu$, and let $u = k'x$ be a control law that minimizes a quadratic performance index and stabilizes the system. Suppose, however, that this control law is not used: rather, the law $u = \rho_0 \operatorname{sat} [\mu_0(k'x)/\rho_0]$ is used where $1 \leq \rho_0 < \infty$, $1 \leq \mu_0 < \infty$, and $\operatorname{sat} y = y$ for $|y| \leq 1$, $= +1$ for $y > 1$, and $= -1$ for $y < -1$. Show that for sufficiently small $x(0)$, depending on ρ_0 and μ_0, the closed-loop system is asymptotically stable. (See [6] for a full discussion of this result.)

Problem 6.1-2. Consider the feedback arrangement of Fig. 6.1-1 with K_α, F, and G as earlier described, and $\phi(\cdot)$ constrained as in Eq. (6.1-6). Suppose that an external input $u(\cdot)$ is applied, with $u'(t)u(t) \leq M$ for all t and some constant

M. By taking $V = x'P_\alpha x$ and computing a modified $\dot{V}$ to take into account the fact that $u(\cdot)$ is no longer zero, show that the states remain bounded, and that this bound depends only on M and $x(0)$ and not on the particular $u(\cdot)$ adopted. Discuss the application of this result to the consideration of stability of systems with nonlinear feedback with hysteresis, when the external input is identically zero. (This is a hard problem.)

6.2 RELAY CONTROL SYSTEMS

In this section, we consider some general properties of control systems containing relays. The results will be applied in the next section to consider the effect of placing relays in the control loop of an otherwise optimal system. In a later chapter, the theory will be applied to certain optimal systems where the optimal control law may be realized using linear gain elements together with a relay.

Our discussion here will start with the definition of a relay system that will take into account the notion of chattering. Following this, stability results will be presented. Our initial discussions make no restriction of time invariance or linearity of the nonrelay part of the system, but do restrict consideration to single-input, single-output systems. (Results for multiple-input, multiple-output systems are complex, and not well detailed in the literature.) The stability discussions will be for restricted classes of systems and will serve as background for considering the effect of introducing relays into nominally optimal linear regulators.

To define relay systems and the notion of chattering, we start by considering the scheme of Fig. 6.2-1, with the equation

$$\dot{x} = f[x, \operatorname{sgn}(h'x), t] \tag{6.2-1}$$

where

$$\begin{aligned} \operatorname{sgn}\sigma(t) &= +1 \quad \text{for } \sigma(t) > 0, \quad \text{or } \sigma(t) = 0 \quad \text{with } \sigma(t-\epsilon) > 0 \\ &= -1 \quad \text{for } \sigma(t) < 0, \quad \text{or } \sigma(t) = 0 \quad \text{with } \sigma(t-\epsilon) < 0 \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{6.2-2}$$

and ϵ is an arbitrarily small positive number.

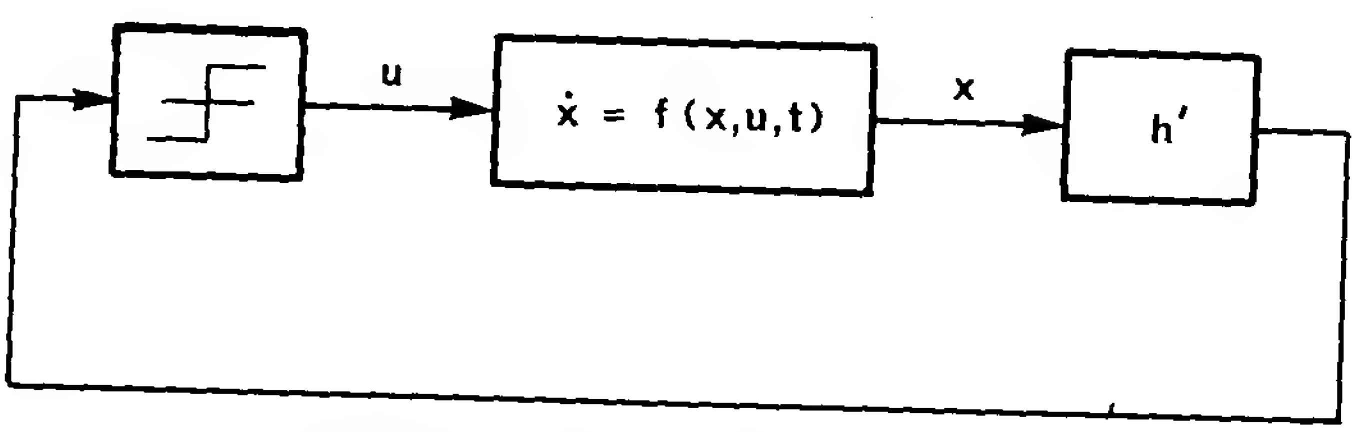


Fig. 6.2-1 Basic relay feedback system.

At this point, several comments are in order. First, we could have considered a more general relay scheme than that of (6.2-1), where $h'x$ is replaced by $h(x, t)$—a scalar, but not necessarily linear, function of x. The analysis would be almost the same. For our purposes, though, the slightly restricted form of (6.2-1) is more convenient. Second, we shall subsequently take specializations of (6.2-1), requiring it to be a linear time-invariant system, save for the relay, and to be a linear time-invariant system, save for input nonlinearity and the relay. There is little point in making the specialization at this stage, however, which we postpone until just before a discussion of stability. Third, the definition of (6.2-1) and (6.2-2), which is due to André and Seibert [8], may seem unreasonably complex. In particular, it may be felt that it would be sufficient to always define sgn 0 as 0, rather than sometimes -1, $+1$, or 0, as in (6.2-2). However, this is not so, and indeed (6.2-2) as it stands even requires special interpretation to avoid certain difficulties, which we shall describe shortly.

Referring to (6.2-1), we see that the switching surface or hyperplane $h'x = 0$ (which is a line where x is a 2 vector, a plane when x is a 3 vector, etc.) divides the state space into two regions, in one of which the control to the system $\dot{x} = f(x, u, t)$ is $+1$, in the other -1. The scalar quantity $\sigma = h'x$ measures the distance of a point x from the nearest point on the switching surface, with due attention paid to algebraic sign, whereas the scalar $\dot{\sigma}$ measures the rate at which an arbitrary point x on a trajectory is approaching the switching surface.

Consider an arbitrary point w on the switching surface, together with the values of σ and $\dot{\sigma}$ near w. Assuming for the moment $\dot{\sigma}$ is nonzero, several situations arise in respect to the possible signs of $\dot{\sigma}$ and σ. In the $\sigma > 0$ region, $\dot{\sigma}$ may be positive or negative, and likewise in the $\sigma < 0$ region. Thus, any of the possibilities shown in Figs. 6.2-2(a) through 6.2-2(d) may hold. These figures show typical trajectories on each side of the switching surface; note that these trajectories have directions consistent with the signs of $\dot{\sigma}$.

In cases (a) and (d) of Fig. 6.2-2, switching is instantaneous. Trajectories approach the switching surface from one side, cross it, and move away from it on the other side. In case (*b*), no trajectories approach the switching surface, and thus no switching occurs. However, if a system is started in a state w on the switching surface, the subsequent trajectory is not uniquely defined, as two trajectories leave w, extending into the $\sigma > 0$ and $\sigma < 0$ regions. In case (c), trajectories in both the $\sigma > 0$ and $\sigma < 0$ regions will head toward the switching surface. No trajectories can leave the point w, and thus the differential equation (6.2-1) apparently cannot be solved beyond this point. For this reason, points such as w are termed "endpoints."

This raises an obvious difficulty: If in Fig. 6.2-2(c), a trajectory reaches the point w, mathematically the system behavior cannot be defined beyond

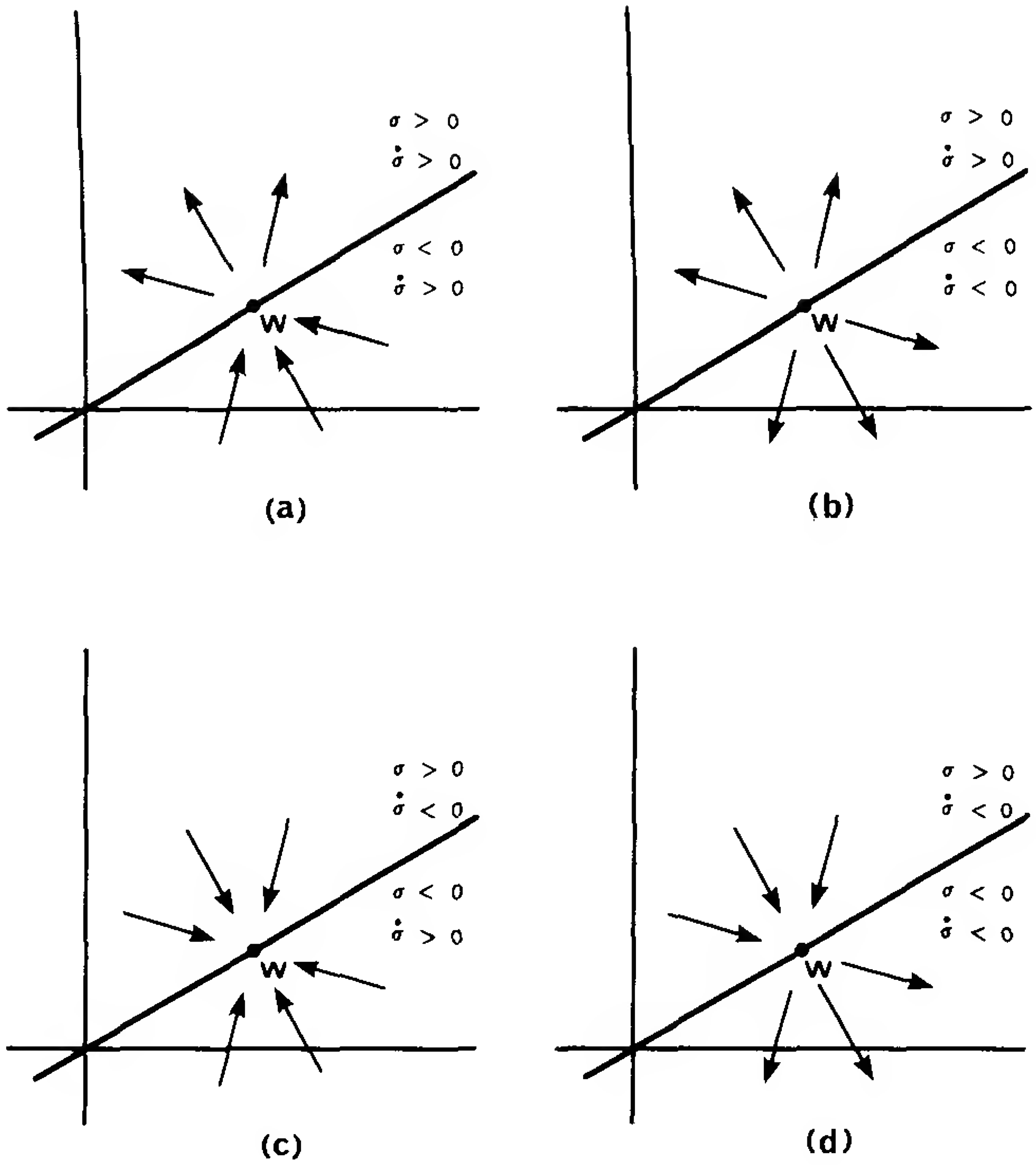


Fig. 6.2-2 Possible trajectories at switching time.

the time when w is reached. Physically, of course, something must happen, and thus we are faced with a paradox. This has been resolved in [8], where one or both of the following variations in relay operation are assumed to take place.

1. Operation of the relay is delayed by a very short time τ. Then if σ changes instantaneously from a positive to a negative value at time $t = 0$, the relay output changes from $+1$ to -1 at time $t = \tau$.
2. There is hysteresis in the relay. It operates when σ has changed sign and passed to a slightly nonzero value, $\epsilon > 0$ for a sign change from negative to positive, and $-\epsilon < 0$ for a sign change from positive to negative.

The effects of (1) and (2) are similar. Supposing that mechanism (1) is present, we see that a trajectory heading for w will continue on past w to the opposite side of the switching surface. At time τ, after crossing the switching surface, the sign of the control will change, causing the trajectory to head back toward the switching surface. However, on reaching the switching surface, the trajectory will again carry on across it as the sign of the control again will not change for a time τ. The net result is that the trajectory zigzags

across the switching surface at a rate determined by τ, and the smaller τ is, the higher will be the frequency of crossings of the switching surface.

Under these conditions, the solution of (6.2-1) is said to exhibit a *chattering mode of behavior*, and this mathematically predicted mode of behavior turns out to be in accord with physical observations. Moreover, the theory covers the situation where $\dot{\sigma}$ can be zero at or near endpoints such as w, a situation hitherto excluded.

Let us now give a quantitative analysis. Instead of (6.2-1), we take as the describing equation

$$\dot{x} = f[x(t), \operatorname{sgn} h'x(t-\tau), t]. \tag{6.2-3}$$

Adopt the notation f^+ to denote $f(x, 1, t)$ and f^- to denote $f(x, -1, t)$. We note first a condition for a state $x = w$ on the switching surface $h'x = 0$ at time t to be an endpoint. Referring to Fig. 6.2-2(c), we see that $\sigma > 0$ implies $\dot{\sigma} < 0$ in the vicinity of w if $h'f^+ < 0$; likewise, $\sigma < 0$ implies $\dot{\sigma} > 0$ in the vicinity of w if $h'f^- > 0$. Note that these conditions are only sufficient for w to be an endpoint. If $h'f^+ = 0$, for example, w may or may not be an endpoint. Restricting ourselves to the simpler case, though, we note the *endpoint condition*:

$$h'f^+ < 0 \quad \text{and} \quad h'f^- > 0. \tag{6.2-4}$$

We shall now consider the chattering mode solution of (6.2-3), and we shall study in particular what happens as τ approaches zero.

Suppose that for $t < t_1$, the state lies in the region $h'x < 0$, and that at t_1, we have $x(t_1) = w$—i.e., the trajectory meets the switching surface. Define $x_2 = x(t_1 + \tau)$, which is the value of the state vector when switching of the control takes place from -1 to $+1$, and let $x_3 = x(t_1 + \Delta t)$ be the state vector when the trajectory next intersects the switching surface, where we are implicitly assuming that w is an endpoint.

In the interval $[t_1, t_1 + \tau]$, we have $\operatorname{sgn} h'x(t-\tau) = -1$, so that

$$x_2 = w + \tau f^- + O(\tau^2).$$

For $t > t_1 + \tau$, the control is $u = +1$. Hence,

$$\begin{aligned} x_3 &= x_2 + (\Delta t - \tau)f^+ + O[(\Delta t - \tau)^2] \\ &= w + \Delta t f^+ + \tau(f^- - f^+) + O(\tau^2) + O[(\Delta t - \tau)^2]. \end{aligned}$$

Multiplying this equation on the left by h' and recalling that x_3 and w are both on the switching surface $h'x = 0$, we obtain

$$\Delta t = -\tau \frac{h'(f^- - f^+)}{h'f^+}.$$

Furthermore,

$$\frac{x_3 - w}{\Delta t} = f^+ - (f^- - f^+) \frac{h'f^+}{h'(f^- - f^+)}.$$

Clearly, as $\tau \to 0$, the system trajectory approaches arbitarily close to the trajectory defined by

$$\dot{x} = f^+ - (f^- - f^+)\frac{h'f^+}{h'(f^- - f^+)}. \tag{6.2-5}$$

As one might expect, as τ approaches zero, the system trajectories are restricted to leaving the switching surface by a smaller and smaller amount, and in the limit must remain on it. This is borne out by evaluating $h'\dot{x}$ using (6.2-5), which is immediately checked to be zero.

Similar arguments carry through when hysteresis is postulated. Let us summarize the salient points.

1. The relay system (6.2-1) is inadequately described by the definition (6.2-2), because of the possible phenomenon of endpoints.
2. "Endpoints" will occur if the constraint (6.2-4) holds.
3. When the system reaches an endpoint, its physical behavior is one of chattering. The states move according to (6.2-5), and remain on the switching surface $h'x = 0$ when infinitely small time delays in switching or hysteresis in the relay are postulated. If and when (6.2-4) subsequently fails, the system leaves the chattering mode, reverting to its regular mode of operation described by (6.2-1) and (6.2-2).

Linear systems with relay feedback. We shall now consider a specialization of the foregoing results, replacing the system (6.2-1) by

$$\dot{x} = Fx - g \operatorname{sgn}(h'x) \tag{6.2-6}$$

where F, g, and h are all time invariant. Figure 6.2-3 shows the arrangement, with the external input in the figure assumed to be zero.

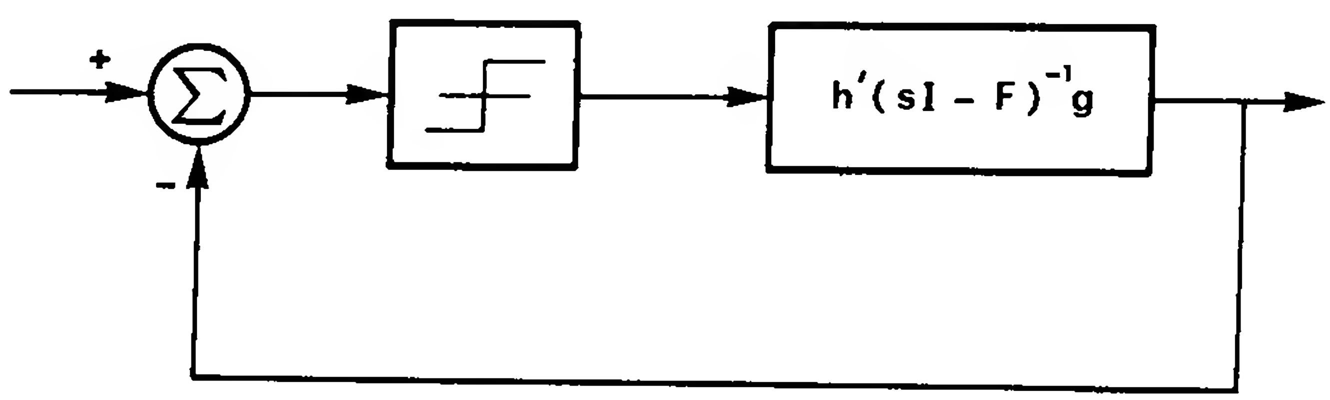


Fig. 6.2-3 System defined by Eq. (6.2-6); the external input is assumed zero.

The functions f^+ and f^- become $Fx - g$ and $Fx + g$. Accordingly, the condition that a point w lying on $h'x = 0$ be an endpoint is, from (6.2-4),

$$-h'g < h'Fw < h'g, \tag{6.2-7}$$

which, in turn, indicates that endpoints occur if and only if

$$h'g > 0. \tag{6.2-8}$$

Motion in the chattering mode is described by (6.2-5) with the appropriate substitutions for f^+ and f^-; the resulting equation is

$$\dot{x} = \left(I - \frac{gh'}{h'g}\right)Fx. \tag{6.2-9}$$

Summing up, Eq. (6.2-9) and the constraint $h'x = 0$ completely describe motion in the chattering mode, with (6.2-7) defining those states on $h'x = 0$ which are endpoints.

A plausible but nonrigorous argument suggesting (6.2-9) is as follows. With $h'x \equiv 0$, it follows that $h'\dot{x} \equiv 0$ or $h'Fx - h'g \operatorname{sgn}(h'x) = 0$. That is, $\operatorname{sgn}(h'x)$ may be replaced by $h'Fx/h'g$. The equation $\dot{x} = Fx - g \operatorname{sgn}(h'x)$ then becomes precisely (6.2-9).

The chattering associated with endpoints could be physically destructive of a relay, and therefore it might be thought that endpoints should be avoided if possible (e.g., by changing the sign of a feedback gain). But it turns out that if $h'g < 0$, implying that there are no endpoints, the resulting closed-loop system is always unstable. (Moreover, there exist states w on the switching surface such that two trajectories emanate from w, and thus the behavior of the system is not completely specifiable.) The remaining possibility, viz., $h'g = 0$, can arise physically. Here again, though, there must be either chattering or instability. To avoid complication, we shall not pursue this condition here.

To avoid chattering, a possible approach is to arrange a dual-mode control system: By monitoring $h'x$ and observing when chattering starts, the control $u = -\operatorname{sgn}[h'x]$ can be replaced by $u = -(h'F/h'g)x$, which is a linear control law. The system then behaves according to (6.2-9). The difficulty that arises with this arrangement is that if, owing to inaccuracies or disturbances, $h'x = 0$ fails and becomes, say, $h'x = \delta$ for some small δ, (6.2-9) does not have the property that $h'x$ will subsequently tend to zero. In other words, asymptotic stability is impossible with this arrangement.

However, there is an attractive modification. Equation (6.2-9) in the first instance represents behavior on the switching surface $h'x = 0$. Therefore, an alternative equation that constitutes an equivalent representation on the switching surface is

$$\dot{x} = \left(I - \frac{gh'}{h'g}\right)Fx - \alpha\frac{gh'}{h'g}x \tag{6.2-10}$$

(where α is an arbitrary constant), since the additional term is identically zero on the switching surface. Now suppose that a dual-mode control is used, based on (6.2-10) rather than (6.2-9). This would require a feedback of $u = -[h'(F + \alpha I)/h'g]x$. As remarked, motion on the switching surface is unaltered, but motion in the $\sigma = h'x$ coordinate direction is affected. Multiplying (6.2-10) on the left by h', it follows that

$$\frac{d}{dt}(h'x) = -\alpha(h'x). \tag{6.2-11}$$

This equation implies that α can be selected to ensure any degree of stability of the σ coordinate if a dual-mode scheme is implemented. This argument is independent of the stability of the chattering mode motion on the switching surface $h'x = 0$.

Introduction of input nonlinearity. As a second specialization of the general results, we replace the system (6.2-1) by

$$\dot{x} = Fx - g\beta[\operatorname{sgn}(h'x), t], \tag{6.2-12}$$

where β is a (possibly time-varying) nonlinearity, with $\beta(1, t)$ and $\beta(-1, t)$ bounded and continuous, and

$$\beta^+(t) = \beta(1, t) \geq \beta_1 > 0 \quad \text{and} \quad \beta^-(t) = \beta(-1, t) \leq \beta_2 < 0 \tag{6.2-13}$$

for all t and certain constants β_1 and β_2. The matrix F and vector g are constant. Figure 6.2-4 shows the arrangement considered, with the external input in the figure assumed zero. The block labeled β is meant to symbolize the presence of nonlinearities, possibly time varying, in the input transducers of the linear system of transfer function $h'(sI - F)^{-1}g$.

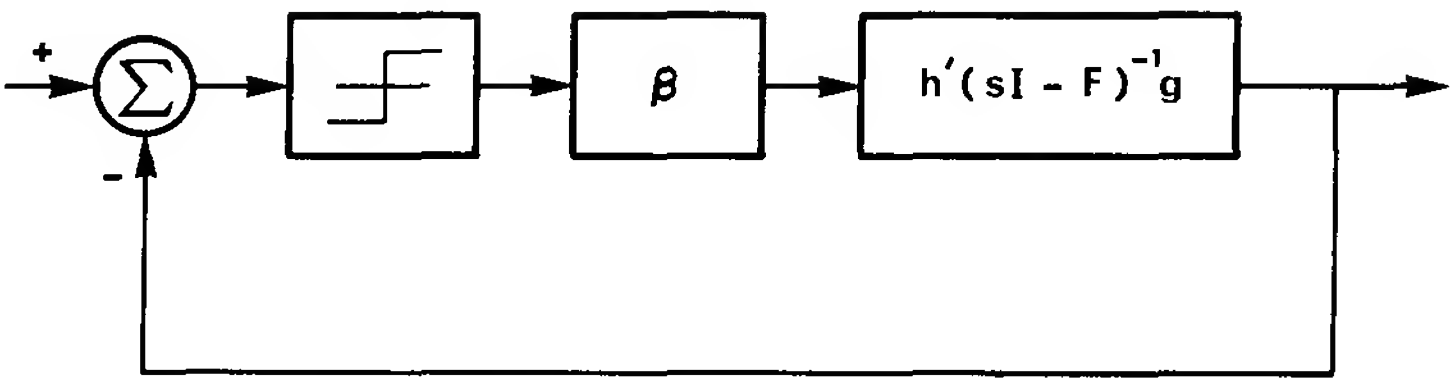


Fig. 6.2-4 Introduction of further nonlinearity into scheme of **Fig. 2-3.**

Functions f^+ and f^- become $Fx - g\beta^+$ and $Fx + g\beta^-$, and a point w lying on $h'x = 0$ is an endpoint if

$$-h'g\beta^- < h'Fw < h'g\beta^+. \tag{6.2-14}$$

Motion in the chattering mode is again described by

$$\dot{x} = \left(I - \frac{gh'}{h'g}\right)Fx. \tag{6.2-9}$$

Consequently, the effect of the relay in the chattering mode is to cancel out the detailed effect of the nonlinearity. The actual chattering mode behavior is unaltered, being still governed by a linear equation, but the set of endpoints is affected. Motion outside the chattering mode is affected; effectively, controls of β^+ and β^- are used in lieu of $+1$ and -1.

The use of a dual-mode controller, feeding back $u = -h'Fx/h'g$ or even $-[h'(F + \alpha I)/h'g]x$ when chattering is detected, is no longer acceptable. Such a linear feedback law will not result in the closed-loop system being linear because of the presence of the β nonlinearity.

Stability. In the remainder of this section, we shall discuss the asymptotic stability of the closed-loop schemes (6.2-6) and (6.2-12), restricting attention first to (6.2-6). For the system trajectories to be well defined, we shall assume that the system, on reaching an endpoint, enters the chattering mode, as described. It is necessary to distinguish in our discussion the concepts of asymptotic stability (which demands boundedness and convergence of the trajectories to zero for a limited set of initial conditions) and global asymptotic stability (which demands boundedness and convergence of the trajectories to zero for any initial condition). We shall present conditions for both asymptotic and global asymptotic stability, beginning with the former.

Both frequency domain and time domain conditions for asymptotic stability can be presented. Although the two types of conditions are equivalent, the equivalence is certainly not obvious a priori, and this fact motivates the presentation of both sets of conditions. In the problems at the end of this section, the student is asked for illumination of the interrelation between the two sets of conditions.

In [9], frequency domain conditions for asymptotic stability are stated. With F of dimension $n \times n$, the transfer function $h'(sI - F)^{-1}g$ can be expressed in the form

$$h'(sI - F)^{-1}g = \frac{b_n s^{n-1} + \cdots + b_1}{s^n + a_n s^{n-1} + \cdots + a_1}. \tag{6.2-15}$$

Then *necessary* conditions for asymptotic stability in the vicinity of the origin are either (1) $b_n > 0$ and the zeros of $b_n s^{n-1} + \cdots + b_1$ have nonpositive real parts, or (2) $b_n = 0$, $b_{n-1} > 0$, $a_n \geq b_{n-2}/b_{n-1}$, and the zeros of $b_n s^{n-1} + \cdots + b_1$ have nonpositive real parts.

Sufficient conditions for asymptotic stability in the vicinity of the origin are either (1) $b_n > 0$ and the zeros of $b_n s^{n-1} + \cdots + b_1$ have negative real parts, or (2) $b_n = 0$, $b_{n-1} > 0$, $a_n > b_{n-2}/b_{n-1}$, and the zeros of $b_n s^{n-1} + \cdots + b_1$ have negative real parts.

Finally, if $b_n = b_{n-1} = 0$, instability is present.

The sufficiency conditions provide a revealing interpretation using root locus ideas. Referring to the system shown in Fig. 6.2-5, consider those points on the root locus corresponding to letting β approach ∞. It is well known that there are n of these points such that m approach zeros of $h'(sI - F)^{-1}g$

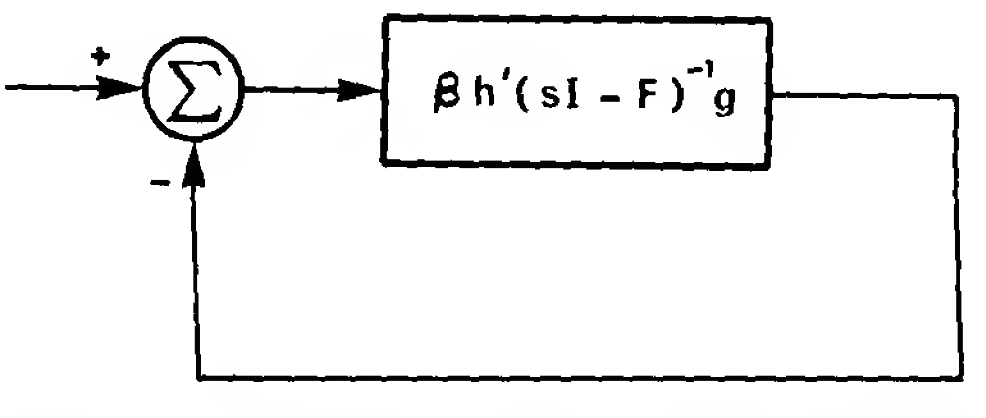


Fig. 6.2-5 Closed-loop system with gain element β.

(where m is the degree of the numerator polynomial of this transfer function), and $n - m$ approach infinity. Moreover, these $(n - m)$ points approach infinity along asymptotes, which, if $n - m \geq 3$, extend into the half-plane $\text{Re}[s] > 0$. If $n - m = 1$, the single asymptote extends along the negative real axis [corresponding to $b_n > 0$ in (6.2-15)]; whereas if $n - m = 2$ [corresponding to $b_n = 0$, $b_{n-1} > 0$ in (6.2-15)], the two asymptotes extend in a direction parallel to the imaginary axis, emanating from the point $-[a_n - (b_{n-2}/b_{n-1})] + j0$, which is evidently on the negative real axis. If either of the two sufficency conditions for asymptotic stability are fulfilled, the system of Fig. 6.2-5 becomes asymptotically stable for suitably large β, and, of course, remains asymptotically stable for further increase in β. The relay result might then have been conjectured by observing that the relay characteristic may be regarded as the limit of a characteristic of the form of Fig. 6.2-6, as β approaches infinity.

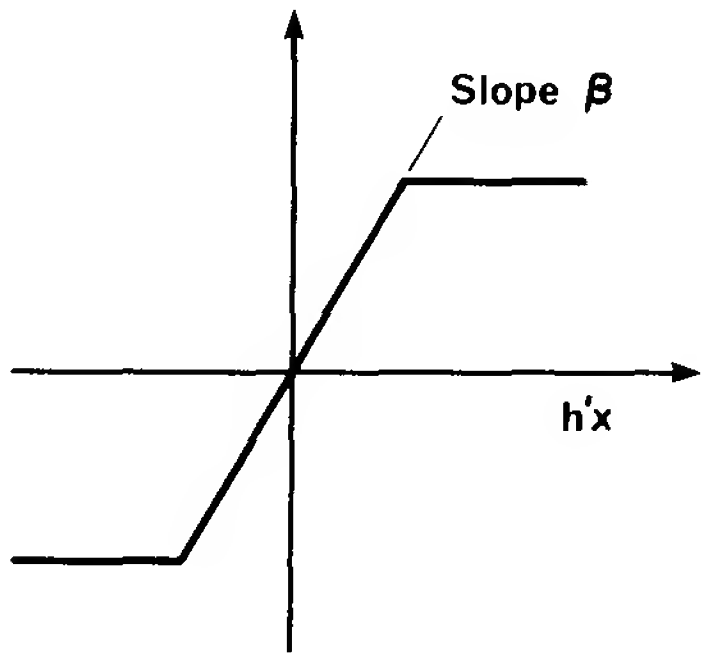


Fig. 6.2-6 One form of approximation to a relay.

A discussion of the time domain approach may be found in, e.g., [10] and [11]. First, consider the motion of (6.2-6) when it starts off in the chattering mode—i.e., when $h'g > 0$, and the initial state x_0 satisfies $|h'Fx_0| < h'g$ [see Eqs. (6.2-7) and (6.2-8)]. Then it is known that the system continues in the chattering mode, at least until the inequality $|h'Fx| < h'g$ fails. Now, if the system is asymptotically stable, it must have the property that for suitably small x_0, the inequality $|h'Fx| < h'g$ always holds for all states x on the system trajectory. Consequently, for suitably small x_0 for which $h'x_0 = 0$, the system if asymptotically stable for all time moves according to

$$\dot{x} = \left(I - \frac{gh'}{h'g}\right)Fx. \tag{6.2-9}$$

Thus, the linear equation (6.2-9) must define an asymptotically stable system† for all x_0 satisfying $h'x_0 = 0$. It turns out (see Problems 6.2-2 and 6.2-4)

†The terminology is a little nonstandard. The normal definition of asymptotic stability demands decay to zero from any initial state near an x_0 satisfying $h'x_0 = 0$. We imply here decay to zero from precisely those initial states x_0 satisfying $h'x_0 = 0$.

that a necessary and sufficient condition for this is that the matrix $[I-(gh'/h'g)]F$ have $(n-1)$ eigenvalues with negative real parts. {The remaining eigenvalue has to be zero, as may be seen from the equation $h'[I-(gh'/h'g)]F = 0$.}

In summary, for any initial condition on the switching surface $h'x = 0$ and suitably close to the origin, the chattering mode trajectories of (6.2-6) are asymptotically stable, provided (6.2-9) is asymptotically stable for any such initial condition. A necessary and sufficient condition for this situation is that $[I-(gh'/h'g)]F$ has $(n-1)$ eigenvalues with negative real parts.

What happens now for initial conditions close to the origin, but not on the switching surface $h'x = 0$? It turns out (see, e.g., [11] and Problem 6.2-5) that so long as asymptotic stability is guaranteed for initial states on the switching surface, it is guaranteed for all suitably small initial states, on or off the switching surface. Consequently, sufficient conditions for asymptotic stability are as follows:

1. $h'g > 0$.
2. $(n-1)$ eigenvalues of $[I-(gh'/h'g)]F$ have negative real parts.

It is also possible to derive time domain asymptotic stability conditions applying for the case $h'g = 0$, but we shall not bother with these here. They are actually the restatement of the frequency domain conditions given earlier for the case when $b_n = 0$ in Eq. (6.2-15). As earlier remarked, the case $h'g < 0$ can be shown to lead to instability (see Problem 6.2-3).

We now make some brief remarks concerning the system (6.2-12), which possesses additional nonlinearity over the system (6.2-6). The equation describing chattering mode motion is the same for (6.2-12) as for (6.2-6); therefore, for those initial states of (6.2-12) for which chattering will subsequently always occur, asymptotic stability conditions are the same as for (6.2-6). It is also not difficult to show that for states of (6.2-12) which are suitably close to but not on that part of the switching surface where states are known to be asymptotically stable, asymptotic stability also holds. Therefore, conditions for asymptotic stability of (6.2-12) are identical with conditions for asymptotic stability of (6.2-6); the precise set of initial states that define asymptotically stable trajectories will, in general, differ for the two systems.

The question of global asymptotic stability is yet more complicated. We discuss it briefly here for the sake of completeness. Note that whereas the asymptotic stability results may be applied to optimal systems, the global asymptotic stability results may not generally be so applied—the next section will discuss these points more fully.

One way to ensure global asymptotic stability is to require that the following set of conditions, associated with the name of Popov, [12], be satisfied.

1. All eigenvalues of F have negative real parts.

2. There exist nonnegative numbers a and b, not both zero, such that $\text{Re}[(j\omega a + b)h'(j\omega I - F)^{-1}g] > 0$ for all real ω, except that if $b = 0$, the inequality sign may be replaced by an equality sign at $\omega = 0$.

A graphical procedure exists for checking the existence of a and b such that the inequality holds. Observe that

$$\begin{aligned}&\text{Re}[(j\omega a + b)h'(j\omega I - F)^{-1}g]\\ &\qquad = b\,\text{Re}[h'(j\omega I - F)^{-1}g] - a\omega Im[h'(j\omega I - F)^{-1}g].\end{aligned}$$

Thus, if a plot is made for varying ω with x coordinate $\text{Re}[h'(j\omega I - F)^{-1}g]$ and y coordinate $\omega Im[h'(j\omega I - F)^{-1}g]$ {as distinct from merely $Im[h'(j\omega I - F)^{-1}g]$, which would be the case for a Nyquist plot}, the inequality condition becomes $bx - ay > 0$ for all points $[x, y]$ on the plot. Therefore, if the plot lies strictly below and/or to the right of any straight line of slope between zero and infinity (save for a line of slope zero, the plot may and will meet the line at the origin), then global asymptotic stability is established.

The regular Nyquist plot of $h'(j\omega I - F)^{-1}g$ can be used to check the inequality condition if either $a = 0$ or $b = 0$, as inspection of the inequality will readily show. Problem 6.2-6 asks for a partial proof of this result. With $b = 0$, the result is established in [10] by construction of a Lyapunov function.

EXAMPLE. Consider the transfer function $(s + 3)/(s + 1)(s + 2)$, which has the state-space realization

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u,$$

$$y = [3 \quad 1]x.$$

Suppose that a relay is inserted in a feedback loop around a plant with these equations. Observe that $h'g = 1$, guaranteeing by (6.2-8) that endpoints occur. From (6.2-7), we conclude that endpoints are defined by

$$3x_1 + x_2 = 0, \qquad -1 < -2x_1 < 1.$$

This is a finite segment, with the origin as midpoint, of the straight line $3x_1 + x_2 = 0$.

Examination of the frequency domain conditions for stability shows that the sufficiency conditions are fulfilled, since the transfer function numerator, $s + 3$, has degree exactly one less than that of the denominator, and has no nonnegative real part zero.

The equation of motion of the chattering regime is

$$\dot{x} = \left(I - \begin{bmatrix} 0 \\ 1 \end{bmatrix} [3 \quad 1]\right) \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} x = \begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix} x.$$

The eigenvalues of

$$\begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix}$$

are zero and -3, which again implies stability through the time-domain criterion.

Last, we can consider global asymptotic stability. Taking $a = 1$, $b = 1.5$ leads to

$$\operatorname{Re}(j\omega a + b)h'(j\omega I - F)^{-1}g = \frac{\omega^4 + 7\omega^2 + 9}{(\omega^2 + 1)(\omega^2 + 2)},$$

from which it is clear that the criteria for global asymptotic stability are fulfilled.

Problem 6.2-1. Given the transfer function of (6.2-15), it is known that the associated relay system is asymptotically stable so long as $b_n > 0$ and the zeros of $b_n s^{n-1} + \cdots + b_1$ have nonpositive real parts. Show that the condition $b_n > 0$ is equivalent to the condition $h'g > 0$. Show also that the condition that the zeros of $b_n s^{n-1} + \cdots + b_1$ have negative real parts is equivalent to the condition that $(n-1)$ eigenvalues of $[I - (gh'/h'g)]F$ have negative real parts. [*Hint for second part:* Write down a state-space realization for $sh'(sI - F)^{-1}g$, and for the inverse of this transfer function. Study the "F matrix" of the inverse, and compare its properties with those of the numerator polynomial of $sh'(sI - F)^{-1}g$. This problem illustrates the equivalence between part of the frequency domain condition and the time domain condition for asymptotic stability.]

Problem 6.2-2. Show that a necessary and sufficient condition for the asymptotic stability of all trajectories of $\dot{x} = [I - (gh'/h'g)]Fx$ with initial condition $h'x_0 = 0$ is that $(n-1)$ eigenvalues of $[I - (gh'/h'g)]F$ have negative real parts.

Problem 6.2-3. Show that if $h'g < 0$, neither of the necessary frequency domain conditions for asymptotic stability holds.

Problem 6.2-4. Suppose that $x = [I - (gh'/h'g)]Fx$ has $n - 1$ eigenvalues with negative real parts. By changing the coordinate basis so that the new F matrix becomes the direct sum of an $(n-1) \times (n-1)$ matrix F_1 and the 1×1 zero matrix, show that there exist nonnegative definite matrices P such that $x'Px \neq 0$ unless $x = \lambda h$ for some constant λ, and such that $V(x) = x'Px$ is a Lyapunov function for all initial states x_0 satisfying $h'x_0 = 0$.

Problem 6.2-5. Let a P matrix be found as in Problem 6.2-4. Consider the tentative Lyapunov function $x'Px + h'x \operatorname{sgn}(h'x)$ for the relay system, as distinct from the linear system $\dot{x} = [I - (gh'/h'g)]Fx$. For all x_0 on $h'x_0 = 0$ and such that the system operates in the chattering mode, this tentative Lyapunov function becomes $x'Px$, and it establishes asymptotic stability for these states. Show also that for x_0 suitably small, but *not* satisfying $h'x_0 = 0$, asymptotic stability still prevails. (Thus it is established that asymptotic stability of the chattering regime alone implies asymptotic stability, with no qualification.)

Problem 6.2-6. Suppose the Popov conditions hold. Except for a pathological situation, they can be shown to imply the existence of a positive definite symmetric

matrix P and a vector l such that $PF + F'P = -ll'$ and $Pg = bh + aF'b + \sqrt{2ah'g}\, l$. Adopt as a Lyapunov function $V = x'Px + 2ah'x \operatorname{sgn}[h'x]$. Assume $h'g > 0$, and show that $\dot{V} \leq 0$ off and on the switching surface. (It is then possible to show that $\dot{V}$ cannot be identically zero along a trajectory, but this is a good deal harder. Consideration of the pathological situation and the case $h'g = 0$ can be made, and global asymptotic stability can then be established.)

6.3 INTRODUCTION OF RELAYS INTO NOMINALLY OPTIMAL SYSTEMS

Our purpose here is to indicate how the results of the preceding section, particularly those on stability, may be put to use in connection with optimal linear systems. Let us assume that a linear system $\dot{x} = Fx + gu$, which is completely controllable, has associated with it an optimal feedback law $u = k'_\alpha x$, resulting from minimizing a performance index of the form $\int_{t_0}^{\infty} e^{2\alpha t}(u^2 + x'Qx)\, dt$. In this index, α is nonnegative and the usual observability requirement is satisfied, guaranteeing that the closed-loop system has degree of stability α. Assume now that the resulting optimal system is modified by the introduction of the relay as indicated in Fig. 6.3-1.

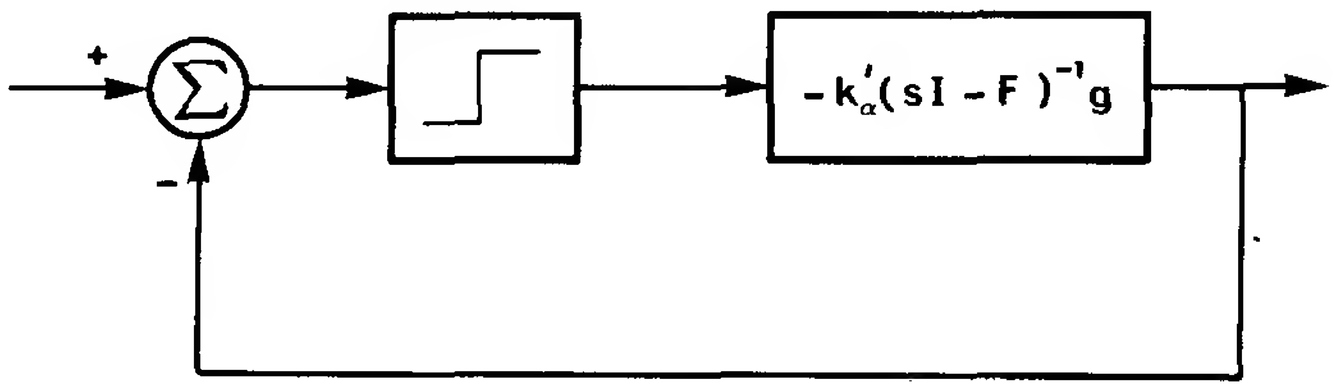


Fig. 6.3-1 Optimal system with relay added in loop.

To carry out an analysis of the system of Fig. 6.3-1, we need to make use of the two equations defining the optimal performance index $x'P_\alpha x$ and optimal control law $u = k'_\alpha x$. These are

$$P_\alpha(F + \alpha I) + (F' + \alpha I)P_\alpha - P_\alpha gg'P_\alpha + Q = 0 \tag{6.3-1}$$

and

$$k_\alpha = -P_\alpha g. \tag{6.3-2}$$

Recall, too, that P_α is that particular solution of $X(F + \alpha I) + (F' + \alpha I)X - Xgg'X + Q = 0$ which is positive definite.

We shall consider first the asymptotic stability of the scheme of Fig. 6.3-1, using frequency domain and then time domain ideas, and subsequently make some brief remarks concerning the global asymptotic stability of the scheme. We shall not consider multiple-input systems or time-varying sys-

tems, although certainly the results would be expected to extend to such systems.

Reference to the preceding section shows that the asymptotic stability properties can be inferred from properties of the transfer function $-k'_\alpha(sI - F)^{-1}g = g'P_\alpha(sI - F)^{-1}g$. In particular, we recall that if the degree of the numerator of the transfer function is one less than that of the denominator, and the leading coefficient of the numerator is positive, a necessary condition for asymptotic stability is that the numerator zeros have nonpositive real parts, and a sufficient condition for asymptotic stability is that the numerator zeros have negative real parts.

We shall now check that the numerator of $g'P_\alpha(sI - F)^{-1}g$ has degree precisely one less than the denominator, with positive leading coefficient. Since the denominator of this transfer function has degree n, it follows that the coefficient of s^{n-1} in the numerator of the transfer function is $\lim_{s\to\infty} sg'P_\alpha(sI - F)^{-1}g$. (If this were zero, there would be no term involving s^{n-1} in the numerator.) Now,

$$\begin{aligned} sg'P_\alpha(sI - F)^{-1}g &= g'P_\alpha(sI - F)(sI - F)^{-1}g + g'P_\alpha F(sI - F)^{-1}g \\ &= g'P_\alpha g + g'P_\alpha F(sI - F)^{-1}g. \end{aligned}$$

Therefore,

$$\lim_{s\to\infty} sg'P_\alpha(sI - F)^{-1}g = g'P_\alpha g, \tag{6.3-3}$$

which is positive because P_α is positive definite.

What of the zeros of $-k'_\alpha(sI - F)^{-1}g$? We recall that the closed-loop optimal system (no relay being present) retains degree of stability α even if a gain constant β is introduced in the feedback loop, for any β in the range $\frac{1}{2} < \beta < \infty$. As β approaches infinity, those poles of the closed-loop system that remain finite must approach the zeros of $-k'_\alpha(sI - F)^{-1}g$, by root locus theory. Now for $\beta < \infty$, the closed-loop poles must lie to the left of $\text{Re}[s] = -\alpha$; therefore, we conclude that the zeros of $-k'_\alpha(sI - F)^{-1}g$ must lie to the left of $\text{Re}[s] = -\alpha$, *or on this line*.

Consequently, if $\alpha = 0$, the zeros of $-k'_\alpha(sI - F)^{-1}g$ have nonpositive real parts, and the necessary (but not sufficient) condition for asymptotic stability is fulfilled, whereas if $\alpha > 0$, the zeros of $-k'_\alpha(sI - F)^{-1}g$ have negative real parts, and the sufficient condition for asymptotic stability is fulfilled.

It may be thought that, for the case $\alpha = 0$, the zeros of $-k'_\alpha(sI - F)^{-1}g$ should have negative, as distinct from nonpositive, real parts. That this need not be so is established from a simple example:

$$F = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad g = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

This case leads to

$$k_0 = \begin{bmatrix} 0 \\ 1 - \sqrt{2} \end{bmatrix},$$

and to a transfer function $-k_0'(sI - F)^{-1}g$, which has a zero for $s = 0$.

We shall now take a time domain approach to discussing the stability of the arrangement of Fig. 6.3-1. In particular, we shall exhibit a Lyapunov function. From the material in the previous section, we know that there will be a chattering regime, since $-k_\alpha' g = g'P_\alpha g > 0$, and that the motion of the system in the chattering regime is governed by

$$\dot{x} = \left[I - \frac{gg'P_\alpha}{g'P_\alpha g}\right] Fx. \tag{6.3-4}$$

For motion on and off the switching surface $g'P_\alpha x = 0$, we adopt as a Lyapunov function

$$V(x) = x'P_\alpha x + g'P_\alpha x \operatorname{sgn}(g'P_\alpha x). \tag{6.3-5}$$

Observe that V is a continuous function of x, on and off the switching surface. It turns out that the function V has sufficient differentiability properties to allow evaluation of $\dot{V}$ except at isolated and therefore insignificant points, located where the system changes from the nonchattering to the chattering mode, or vice versa.

First, off the switching surface, we have

$$\dot{x} = Fx - g \operatorname{sgn}(g'P_\alpha x), \tag{6.3-6}$$

and so

$$\begin{aligned}\dot{V} = x'(P_\alpha F + F'P_\alpha)x &- 2x'P_\alpha g \operatorname{sgn}(g'P_\alpha x) + g'P_\alpha Fx \operatorname{sgn}(g'P_\alpha x) \\ &- g'P_\alpha g[\operatorname{sgn}(g'P_\alpha x)]^2.\end{aligned}$$

Now use (6.3-1) to obtain

$$\dot{V} = -x'(2\alpha P_\alpha + Q - P_\alpha gg'P_\alpha)x - x'(F'P_\alpha g + 2P_\alpha g) \operatorname{sgn}(g'P_\alpha x) - g'P_\alpha g. \tag{6.3-7}$$

That is, $\dot{V}$ is the sum of terms that are quadratic in x, linear in x, and constant. The constant term is always negative; therefore, for suitably small x, $\dot{V} < 0$. Moreover, since $V \to 0$ as $x'x \to 0$, the ratio $-\dot{V}/V$ increases without bound, implying that the instantaneous degree of stability becomes unbounded. On the switching surface, however, a different situation occurs. First, $g'P_\alpha x$ is identically zero, so that, from (6.3-5), $V(x)$ is simply $x'P_\alpha x$. Using (6.3-4), we have

$$\dot{V}(x) = x'(P_\alpha F + F'P_\alpha)x - x'\left(\frac{P_\alpha gg'P_\alpha}{g'P_\alpha g} F + F'\frac{P_\alpha gg'P_\alpha}{g'P_\alpha g}\right)x.$$

But since $g'P_\alpha x$ is zero, the second term disappears. Then, using (6.3-1) and

expanding the first term, we obtain

$$\dot{V}(x) = -2\alpha x'P_\alpha x - x'Qx + x'P_\alpha gg'P_\alpha x.$$

Again, the last term is zero, and so

$$\dot{V}(x) = -2\alpha x'P_\alpha x - x'Qx. \tag{6.3-8}$$

For $\alpha > 0$, asymptotic stability is immediate. Indeed, since now V is simply $x'P_\alpha x$, Eq. (6.3-8) implies $\dot{V}/V \leq -2\alpha$, so that there is a degree of stability equal to α. This is to be expected, because the nonzero eigenvalues of $[I - (gg'P_\alpha/g'P_\alpha g)]F$ (which determine the stability properties in the chattering regime) are none other than the zeros of the numerator polynomial of $-k'_\alpha(sI - F)^{-1}g$, which have been shown to have real part less than or equal to $-\alpha$.

The techniques of the previous chapter also allow the system of Fig. 6.3-2 to be handled, where β is a time-varying nonlinearity with the usual constraints—viz., $\beta(1, t) \geq \beta_1 > 0$ and $\beta(-1, t) \leq \beta_2 < 0$ for all t together with continuity and boundedness of $\beta(1, t)$ and $\beta(-1, t)$. Conditions for asymptotic stability of the system of Figs. 6.3-1 and 6.3-2 are the same. It also turns out that the Lyapunov function of (6.3-5), used in connection with the system of Fig. 6.3-1, serves as a Lyapunov function for the system of Fig. 6.3-2. Verification of this fact is sought in Problem 6.3-3.

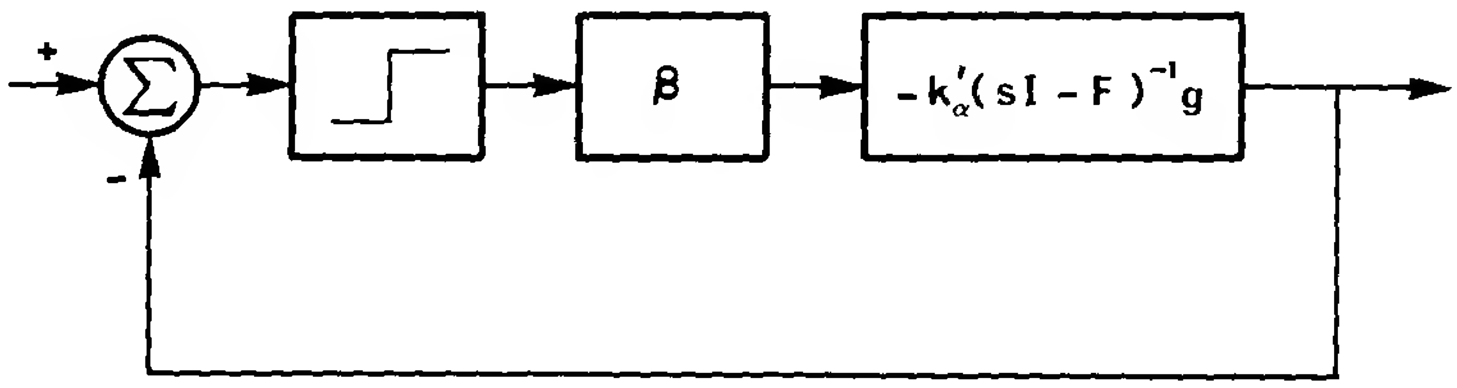


Fig. 6.3-2 Relay and nonlinearity associated with optimal system.

In the first section of this chapter, we demonstrated stability properties for nominally optimal regulators that had input transducer nonlinearities. We did not establish stability of course for arbitrary nonlinearities. However, the insertion of a relay does serve to guarantee asymptotic stability for an almost arbitrary class of input nonlinearities. Therefore, there may be advantages in using a relay other than for preventing saturation. What insertion of the relay unfortunately does not guarantee is global asymptotic stability. Thus, the use of a relay for stabilizing purposes must be with caution.

As remarked earlier, it is not possible to draw general conclusions about the global asymptotic stability of optimal linear systems into which a relay has been inserted, and the Popov criterion of the previous section seems to be the only helpful tool. There is, though, one class of transfer functions $-k'_\alpha(sI - F)^{-1}g$ that will often arise for which global asymptotic stability is

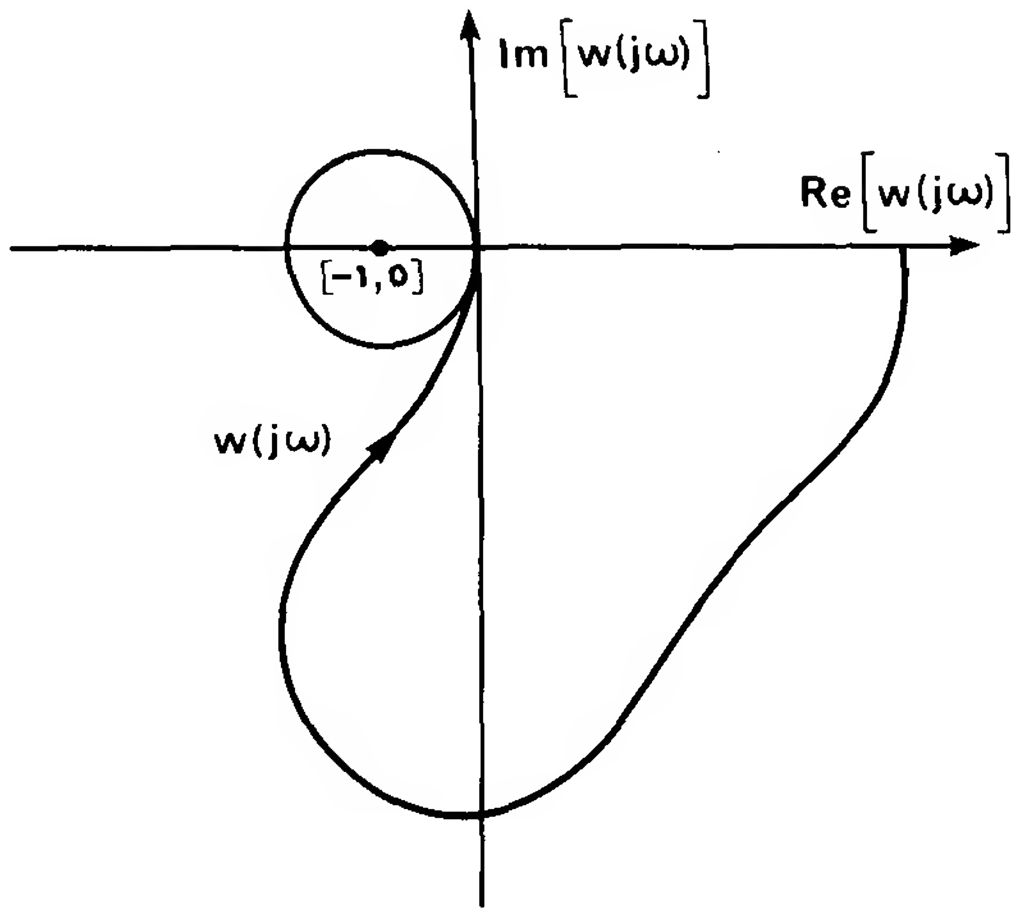


Fig. 6.3-3 Nyquist plot from which global asymptotic stability for system with relay could be concluded.

an immediate consequence of the Popov criterion. This is the class for which F is asymptotically stable and the Nyquist plot of $w(j\omega) = -k'_\alpha(j\omega I - F)^{-1}g$ avoids, for positive ω, the half-plane $\text{Im}[w(j\omega)] \geq 0$ (see Fig. 6.3-3). In this instance,

$$-\omega\,\text{Im}[-k'_\alpha(j\omega I - F)^{-1}g] > 0$$

for all real ω, except for $\omega = 0$, because for positive ω, the Nyquist plot property just described guarantees $\text{Im}[w(j\omega)] < 0$, except for $\omega = 0$. The Popov criterion conditions are then satisfied.

Problem 6.3-1. Consider the system of Fig. 6.3-1, and suppose that a constant gain β is inserted in the feedback part of the loop. Show that the equation of the chattering regime is unaltered. Can you prove that the equation also remains unaltered with time-varying and nonlinear β? Discuss the stability properties of the closed-loop system with the aid of Lyapunov functions. Is it possible to conclude that $\dot{V} \leq 0$ for a larger or smaller set of x when variation from unity feedback is introduced?

Problem 6.3-2. Sketch the graph of the describing function of a relay (viz., the graph of the amplitude of the fundamental component of the relay output with relay input $A \sin \omega t$, versus A). Discuss why this graph suggests asymptotic stability for sufficiently small initial conditions of the closed-loop system of Fig. 6.3-1.

Problem 6.3-3. With notation as defined in this section, show that a Lyapunov function for the system of Fig. 6.3-2 is provided by $x'P_\alpha x + g'P_\alpha x \,\text{sgn}\,(g'P_\alpha x)$.

REFERENCES

[1] Sandberg, I. W., "A Frequency Domain Condition for the Stability of Systems Containing a Single Time-Varying Nonlinear Element," *Bell System Tech. J.*, Vol. 43, No. 4, July 1964, pp. 1601–1638.

[2] Freedman, M., and G. Zames, "Logarithmic Variation Criteria for the Stability of Systems with Time-Varying Gains," *SIAM J. Control*, Vol. 6, No. 3, August 1968, pp. 487–507.

[3] Moore, J. B., and B. D. O. Anderson, "Trade-off Between Loop Gain Time Variations and Relative Stability of Linear Systems," *Proc. Hawaii Intern. Conf. System Sci.*, 1970.

[4] Krasovskii, N. N., *Stability of Motion*, Stanford Press, Stanford, Calif., 1963.

[5] Moore, J. B., and B. D. O. Anderson, "Applications of the Multivariable Popov Criterion," *Intern. J. Control*, Vol. 5, No. 4, 1967, pp. 345–353.

[6] Bass, R. W., and I. Gura, "High Order System Design via State-Space Considerations," *Proc. Joint Auto. Control Conf.*, 1965, R. P. I. Rensselaer, N. Y., pp. 311–318.

[7] Anderson, B. D. O., and J. B. Moore, "Tolerance of Nonlinearities in Time-Varying Optimal Systems," *Electron. Letters*, Vol. 3, No. 6, June 1967, pp. 250–251.

[8] André, J., and P. Seibert, "Über stückweise lineare Differentialgleichungen, die bei Regelungsproblemen auftreten, I and II," *Arch. Math.*, Vol. 7, 1956, pp. 148–156, 157–165.

[9] Anosov, D. V., "Stability of the Equilibrium Position in Relay Systems," *Automation and Remote Control*, Vol. 20, No. 2, February 1959, pp. 130–143.

[10] Alimov, Y. I., "Lyapunov Functions for Relay Control Systems," *Automation and Remote Control*, Vol. 21, No. 6, June 1960, pp. 720–728.

[11] Weissenberger, S., "Stability-Boundary Approximation for Relay-Control Systems with a Steepest Ascent Construction of Lyapunov Functions," *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 88, June 1966, pp. 419–428.

[12] Popov, V. M., "Absolute Stability of Nonlinear Systems of Automatic Control," *Automation and Remote Control*, Vol. 22, No. 8, March 1962, pp. 857–875.

CHAPTER 7

SENSITIVITY ASPECTS OF OPTIMAL REGULATORS

7.1 THE SENSITIVITY PROBLEM

In this section, we shall introduce the so-called sensitivity problem and present a number of background results. The results are then applied in the later sections of this chapter to develop further properties and techniques associated with the optimal regulator.

One of the original motivations for placing feedback around systems is to reduce their sensitivity to parameter variations. Thus, consider the arrangements of Fig. 7.1-1. Both schemes provide the same nominal gain, viz., $A/(1 + A\beta)$. But now if the block with gain A suffers a gain change of 1%, the closed-loop scheme undergoes a gain change of approximately $1/(1 + A\beta)\%$, whereas the open-loop scheme undergoes a gain change of 1%. Consequently, if $1 + A\beta$ is made large, the closed-loop scheme is judged better than the open-loop scheme from the point of view of sensitivity of input-output performance to plant parameter variations [where we think of the block of gain A as the plant, and the blocks of gain β and $1/(1 + A\beta)$ as controllers].

Although this notion applies to the situation where A and β may be frequency dependent, there is no straightforward extension to, say, multiple-input systems, or situations where the varying plant parameter is not gain but something else, such as an initial condition. Therefore, we shall be concerned in this section with setting out on a firm basis what we mean by sensitivity, and sensitivity reduction, in these more general situations.

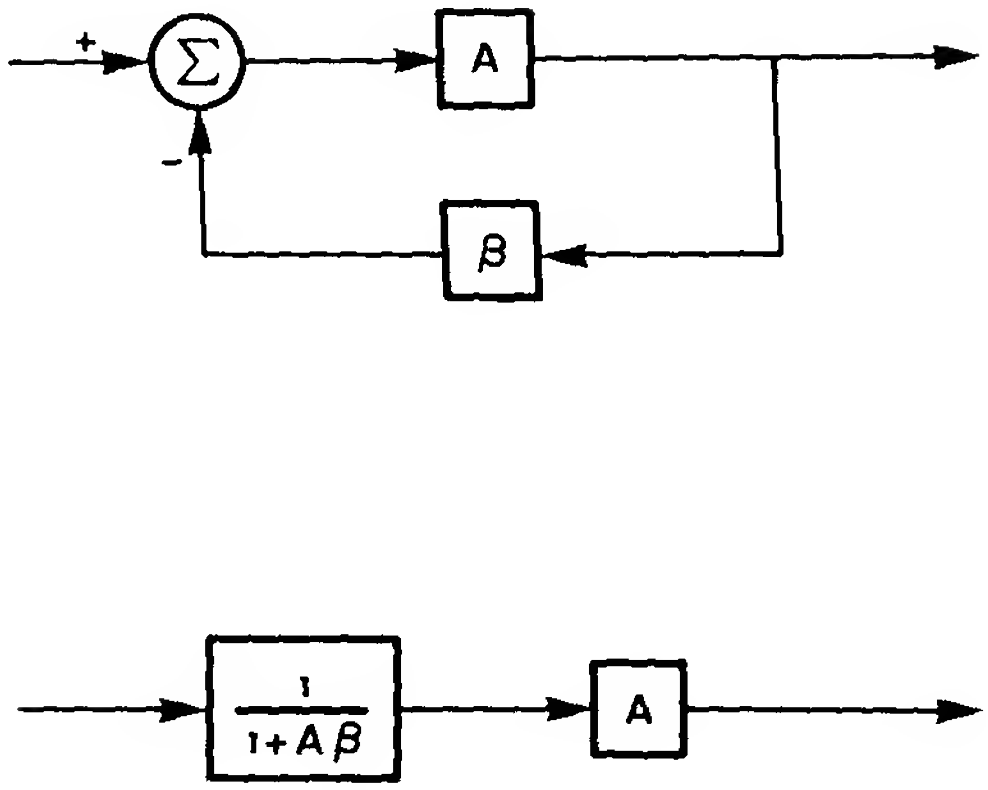


Fig. 7.1-1 A closed-loop scheme and an open-loop scheme.

To begin with, we consider the closed-loop and open-loop schemes of Fig. 7.1-2. Two ways are illustrated of controlling the plant, which is allowed to have multiple inputs and outputs. We also make the following assumptions.

Assumption 7.1-1. The overall transfer matrices of the two schemes—viz., $[I + P(s)F(s)]^{-1}P(s)$ and $P(s)C(s)$—are the same;

Assumption 7.1-2. The controlled systems are asymptotically stable. {In particular, $[I + P(j\omega)F(j\omega)]^{-1}$ must exist for all real ω.}

The effect of these two assumptions is to guarantee that the effect of initial conditions on the output will die out. When this has occurred, identical inputs to the two schemes will produce identical outputs, permitting the per-

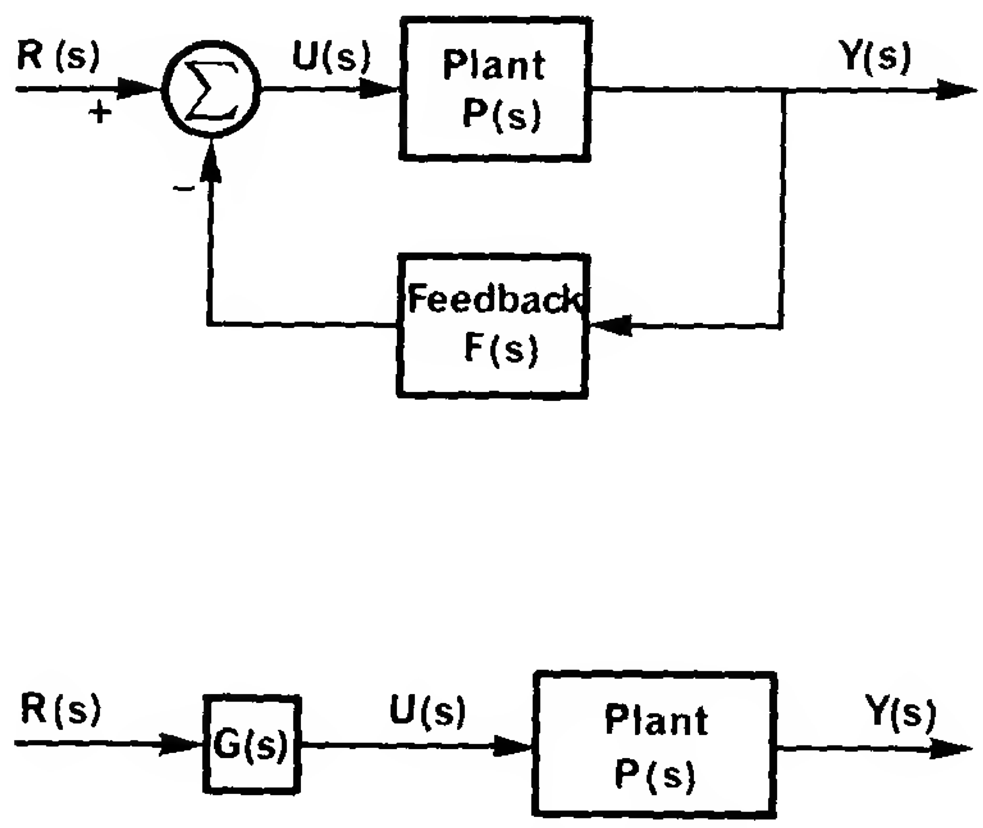


Fig. 7.1-2 Feedback compensator and series compensator.

formance of the two schemes to be meaningfully compared. As will be shown subsequently, the *return difference matrix* $I + P(j\omega)F(j\omega)$, so-called by analogy with the corresponding scalar quantity, will play a key role in quantitatively making the comparison.

We now permit the possibility of parameter variations in the plant. This means that the transfer function matrix of the plant must be regarded as a function of a parameter μ, and that μ can vary away from a nominal value μ_{nom}. Plant parameter variations may also be of the type where that part of the output due to plant initial conditions (and this part alone) may be affected. This would be the case if, e.g., the plant parameter in question was actually an initial condition, or if some states of the plant were uncontrollable but observable, and the way these uncontrollable states were coupled to the output was varied. Also, both the transfer function matrix and the initial condition part of the output may be parameter dependent. Accordingly, we shall write

$$Y(s; \mu) = P(s; \mu)U(s; \mu) + Z(s; \mu). \tag{7.1-1}$$

In this equation, which applies to both the open- and closed-loop arrangements, $Y(s; \mu)$ is the (Laplace transform of the) output of the plant, $P(s; \mu)$ is the plant transfer function matrix, $Z(s; \mu)$ is that component of the plant output due to initial conditions, and $U(s; \mu)$ is the input to the plant (*not* the input to the whole controlled system). Notice that, at least for closed-loop control, $U(s; \mu)$ depends on μ because it partially depends on the feedback output. Normally, $\mu = \mu_{\text{nom}}$, the nominal value of μ.

Now we aim to compare the two schemes of Fig. 7.1-2. Accordingly, assume that the initial conditions on the plant, as well as the external inputs $R(s)$, are the same for both cases. In other words, if the parameter is at its nominal value, the input $U(s; \mu_{\text{nom}})$ to the plant itself is the same for both schemes. When the plant parameter changes, we shall assume it changes in the same way for both plants; accordingly, the variation in the plant output $Y(s; \mu)$ due to initial conditions will be the same in both cases. But the variation in $Y(s; \mu)$ due to the plant inputs will be different. Certainly, $P(s; \mu)$ will change in the same way for both schemes. For the open-loop scheme, $U(s; \mu)$ will be unchanged from $U(s; \mu_{\text{nom}})$; however, for the closed-loop scheme, that component of $U(s; \mu)$ due to $R(s)$ will remain unchanged from the corresponding component of $U(s; \mu_{\text{nom}})$, but that component due to feedback of the plant output $Y(s; \mu)$ will be changed. Thus, there is an overall change in the plant input, which hopefully will compensate for parameter variation.

We now consider these notions quantitatively. Let subscripts o and c refer to the open- and closed-loop schemes, respectively, and to keep the notation as simple as possible, let us omit the Laplace transform variable for the next few equations. We shall be deriving a fundamental equation, (7.1-4), that compares the variation caused in the output by a plant parameter change

for the open- and closed-loop control arrangements. Before we can compare these variations, $Y_o(\mu) - Y_o(\mu_{\text{nom}})$ for the open-loop arrangement and $Y_c(\mu) - Y_c(\mu_{\text{nom}})$ for the closed-loop arrangement, we shall derive separate expressions for both. With a parameter change from μ_{mon} to an arbitrary μ, the external input R is, of course, unaltered. Also, at the nominal parameter value, the plant inputs $U_c(\mu_{\text{nom}})$ and $U_o(\mu_{\text{nom}})$ must be the same—likewise for their outputs.

For the open-loop control scheme, so long as R remains the same when the parameter is varied, we have $U_o(\mu_{\text{nom}}) = U_o(\mu)$. Thus, from Eq. (7.1-1),

$$Y_o(\mu) - Y_o(\mu_{\text{nom}}) = [P(\mu) - P(\mu_{\text{nom}})]U_o(\mu_{\text{nom}}) + Z(\mu) - Z(\mu_{\text{nom}}). \tag{7.1-2}$$

As remarked previously, when μ is varied, the input to the plant for the closed-loop control scheme is varied, even though R is left unaltered. Thus, (7.1-1) yields

$$Y_c(\mu) - Y_c(\mu_{\text{nom}}) = P(\mu)U_c(\mu) - P(\mu_{\text{nom}})U_c(\mu_{\text{nom}}) + Z(\mu) - Z(\mu_{\text{nom}}).$$

We can, however, eliminate $U_c(\mu)$. Reference to Fig. 7.1-2 shows that

$$\begin{aligned} U_c(\mu) &= -FY_c(\mu) + R \\ &= -F[Y_c(\mu) - Y_c(\mu_{\text{nom}})] - FY_c(\mu_{\text{nom}}) + R \\ &= -F[Y_c(\mu) - Y_c(\mu_{\text{nom}})] + U_c(\mu_{\text{nom}}) \end{aligned}$$

where we have implicitly used the fact that R, the controlled system input, is the same for two values of μ. Using this expression for $U_c(\mu)$, we have

$$\begin{aligned} Y_c(\mu) - Y_c(\mu_{\text{nom}}) = &-P(\mu)F[Y_c(\mu) - Y_c(\mu_{\text{nom}})] \\ &+ [P(\mu) - P(\mu_{\text{nom}})]U_c(\mu_{\text{nom}}) \\ &+ Z(\mu) - Z(\mu_{\text{nom}}) \end{aligned}$$

or

$$\begin{aligned} &[I + P(\mu)F][Y_c(\mu) - Y_c(\mu_{\text{nom}})] \\ &\qquad = [P(\mu) - P(\mu_{\text{nom}})]U_c(\mu_{\text{nom}}) + Z(\mu) - Z(\mu_{\text{nom}}). \end{aligned} \tag{7.1-3}$$

Equations (7.1-2) and (7.1-3) are the two expressions for the output variations resulting from plant parameter variation from μ_{nom} to μ. They are, perhaps, of some independent interest for design purposes, but the connection between them is of greater importance. Recalling that $U_c(\mu_{\text{nom}})$ and $U_o(\mu_{\text{nom}})$, the two plant inputs corresponding to the two control configurations, are the same, it follows immediately from (7.1-2) and (7.1-3) that

$$Y_c(s; \mu) - Y_c(s; \mu_{\text{nom}}) = [I + P(s; \mu)F(s)]^{-1}[Y_o(s; \mu) - Y_o(s; \mu_{\text{nom}})]. \tag{7.1-4}$$

Equation (7.1-4) relates the variations in the open- and closed-loop plant outputs. We recall it is valid for the following:

1. All system inputs $R(s)$, i.e., the relation is independent of the particular input used, so long as the same input is applied to the open- and closed-loop systems.
2. Any form of plant parameter variations. This includes variation of the transfer function matrix, initial condition response, multiple parameter variations, or any combination of these.

Of course, the plant and controller transfer function matrices do not have to consist of rational functions of s.

If the variation from μ_{nom} to μ is small, and if $Y_o(s; \mu)$, $Y_c(s; \mu)$, and $P(s; \mu)$ depend continuously on μ, Eq. (7.1-4) may be replaced by the approximate equation

$$Y_c(s; \mu) - Y_c(s; \mu_{\text{nom}}) = [I + P(s; \mu_{\text{nom}})F(s)]^{-1}[Y_o(s; \mu) - Y_o(s; \mu_{\text{nom}})], \tag{7.1-5}$$

where the variations of outputs are related by a formula involving the nominal rather than the perturbed value of the parameter. This equation may therefore be more useful, even though it is an approximation.

As another interpretation of (7.1-5), we can also define *sensitivity functions* $\sigma_c(t)$ and $\sigma_o(t)$ associated with the closed- and open-loop outputs. These are given by

$$\sigma_c(t) = \left[\frac{\partial}{\partial\mu}(y_c(t))\right]_{\mu=\mu_{\text{nom}}} \qquad \sigma_o(t) = \left[\frac{\partial}{\partial\mu}(y_o(t))\right]_{\mu=\mu_{\text{nom}}} \tag{7.1-6}$$

assuming the differentials exist. Then Eq. (7.1-5) implies the following relation between the Laplace transforms $\Sigma_c(s)$ and $\Sigma_o(s)$ of the closed- and open-loop sensitivity functions:

$$\Sigma_c(s) = [I + P(s; \mu_{\text{nom}})F(s)]^{-1}\Sigma_o(s). \tag{7.1-7}$$

For the remainder of this section, we shall discuss the question of characterizing *reduction*, via feedback, of sensitivity to plant parameter variations. We shall first state a criterion for a closed-loop scheme to be better than an open-loop one in terms of variations caused in the plant output of each scheme by a parameter variation. Then we shall interpret this criterion as a constraint on the return difference matrix of the feedback arrangement, with the aid of Eq. (7.1-5).

Suppose that variation of a parameter from the nominal value μ_{nom} to $\mu = \mu_{\text{nom}} + \Delta\mu_{\text{nom}}$, where $\Delta\mu_{\text{nom}}$ is small, causes a change in the closed-loop system output of $y_c(t; \mu) - y_c(t; \mu_{\text{nom}})$ and in the open-loop system output of $y_o(t; \mu) - y_o(t; \mu_{\text{nom}})$. We say that *the closed-loop scheme is superior to the open-loop scheme if for all t_1 and all $y_o(\cdot\,; \mu_{\text{nom}})$,*

$$\begin{aligned}\int_{t_0}^{t_1} &[y_c(t; \mu) - y_c(t; \mu_{\text{nom}})]'[y_c(t; \mu) - y_c(t; \mu_{\text{nom}})]\,dt \\ &\leq \int_{t_0}^{t_1} [y_o(t; \mu) - y_o(t; \mu_{\text{nom}})]'[y_o(t; \mu) - y_o(t; \mu_{\text{nom}})]\,dt.\end{aligned} \tag{7.1-8}$$

(Here, t_0 is an arbitrary initial time.) Equation (7.1-8) is equivalent to requiring that the integral squared error for the closed-loop scheme never exceeds that for the open-loop scheme.

By use of Parseval's theorem and several technical artifices, it is possible to show that (7.1-8) implies

$$\int_{-\infty}^{\infty} [Y_c(j\omega;\mu) - Y_c(j\omega;\mu_{\text{nom}})]^{*\prime}[Y_c(j\omega;\mu) - Y_c(j\omega;\mu_{\text{nom}})]\,d\omega$$
$$\leq \int_{-\infty}^{\infty} [Y_o(j\omega;\mu) - Y_o(j\omega;\mu_{\text{nom}})]^{*\prime}[Y_o(j\omega;\mu) - Y_o(j\omega;\mu_{\text{nom}})]\,d\omega.$$

Now use Eq. (7.1-5) {and recall that $[I + P(j\omega;\mu)F(j\omega)]^{-1}$ has been assumed to exist, to guarantee asymptotic stability of the closed-loop scheme}. We obtain, then,

$$\int_{-\infty}^{\infty} [Y_o(j\omega;\mu) - Y_o(j\omega;\mu_{\text{nom}})]^{*\prime}[I - S^{*\prime}(j\omega)S(j\omega)]$$
$$[Y_o(j\omega;\mu) - Y_o(j\omega;\mu_{\text{nom}})]\,d\omega \geq 0 \tag{7.1-9}$$

where

$$S(s) = [I + P(s;\mu_{\text{nom}})F(s)]^{-1}. \tag{7.1-10}$$

Observe that $S(s)$ is the inverse of the return difference matrix. A sufficient condition for (7.1-9) to always hold is thus

$$I - S^{*\prime}(j\omega)S(j\omega) \geq 0. \tag{7.1-11}$$

Equivalently, with $T(s) = I + P(s;\mu_{\text{nom}})F(s)$ (the return difference matrix), we have

$$T^{*\prime}(j\omega)T(j\omega) - I \geq 0. \tag{7.1-12}$$

Consequently, (7.1-11) *and* (7.1-12) *constitute sufficient conditions for sensitivity reduction using feedback.*

Actually, (7.1-11) and (7.1-12) would be necessary if the variation $Y_o(j\omega;\mu) - Y_o(j\omega;\mu_{\text{nom}})$ were an arbitrary Fourier transform. But it is the Fourier transform of $y_o(t;\mu) - y_o(t;\mu_{\text{nom}})$, which is not in general an arbitrary function. Hence, strictly speaking, (7.1-11) and (7.1-12) are not necessary. Apparently, though, the sorts of variation that arise are such as to make these equations "almost necessary," and one could safely rule out the existence of any other condition on the return difference matrix for sensitivity reduction in most cases.

In the scalar input case, the classical result

$$|T(j\omega)| \geq 1 \tag{7.1-13}$$

is recovered. This is the result we noted at the start of the section, but this derivation has, of course, given the constraint (7.1-13) a much wider interpretation than previously.

There is some validity in the concept of degree of sensitivity improvement (as distinct from sensitivity improvement itself) associated with equivalent closed- and open-loop systems. Thus, consider two control systems, each being realized in both closed- and open-loop form. We could say that system 1 offered a greater degree of sensitivity improvement than system 2, if the inequality

$$\int_{t_0}^{t_1} [y_{c1}(t;\mu) - y_{c1}(t;\mu_{\text{nom}})]'[y_{c1}(t;\mu) - y_{c1}(t;\mu_{\text{nom}})]\,dt$$
$$\leq \int_{t_0}^{t_1} [y_{o1}(t;\mu) - y_{o1}(t;\mu_{\text{nom}})]'[y_{o1}(t;\mu) - y_{o1}(t;\mu_{\text{nom}}]\,dt$$

were in some way "stronger" than the inequality

$$\int_{t_0}^{t_1} [y_{c2}(t;\mu) - y_{c2}(t;\mu_{\text{nom}})]'[y_{c2}(t;\mu) - y_{c2}(t;\mu_{\text{nom}})]\,dt$$
$$\leq \int_{t_0}^{t_1} [y_{o2}(t;\mu) - y_{o2}(t;\mu_{\text{nom}})][y_{o2}(t;\mu) - y_{o2}(t;\mu_{\text{nom}})]\,dt.$$

Equivalently, this would require the inequality

$$I - S_1^{*\prime}(j\omega)S_1(j\omega) \geq 0$$

to be "stronger" than the inequality

$$I - S_2^{*\prime}(j\omega)S_2(j\omega) \geq 0.$$

If S_1 and S_2 are of the same dimension, it is clear that the rigorous interpretation of this statement regarding the "strength" of the two inequalities is

$$I - S_1^{*\prime}(j\omega)S_1(j\omega) \geq I - S_2^{*\prime}(j\omega)S_2(j\omega)$$

or

$$S_2^{*\prime}(j\omega)S_2(j\omega) \geq S_1^{*\prime}(j\omega)S_1(j\omega). \tag{7.1-14}$$

Equivalently, one can require

$$T_1^{*\prime}(j\omega)T_1(j\omega) \geq T_2^{*\prime}(j\omega)T_2(j\omega), \tag{7.1-15}$$

which certainly accords with the well-known ideas pertaining to scalar systems. For the latter, (7.1-15) becomes

$$|T_1(j\omega)| \geq |T_2(j\omega)|. \tag{7.1-16}$$

The interpretation is that larger return differences give greater reduction in sensitivity than smaller return differences.

The results of this section stem from work by Perkins and Cruz, amongst whose papers we note [1]. The approach taken here is closely allied to that of Kreindler [2]. It is also interesting to note that all the ideas may be extended to time-varying systems [3].

Besides the property (7.1-11) asked of $S(s)$—the inverse of the return difference matrix, to ensure sensitivity improvement via feedback, we recall

Assumption 7.1-2, which implies that $S(s) = [I + P(s; \mu_{\text{nom}})F(s)]^{-1}$ should have analytic elements in $\text{Re}[s] \geq 0$. Square matrices possessing these two properties are termed *passive scattering matrices* in the network theory literature, where they play a prominent role—see e.g., [4].

EXAMPLE. Consider a plant with transfer function $1/(s + 1)(s + 10)$ and suppose a feedback controller of transfer function a constant, K, is used. The feedback is negative.

The return difference is the loop gain plus unity, or

$$T(j\omega) = 1 + \frac{K}{(j\omega + 1)(j\omega + 10)}$$
$$= \frac{-\omega^2 + (10 + K) + 11j\omega}{-\omega^2 + 10 + 11j\omega}.$$

It follows that

$$|T(j\omega)|^2 - 1 = \frac{K(-2\omega^2 + 20 + K)}{(-\omega^2 + 10)^2 + 121\omega^2}.$$

We conclude that for no value of K will this quantity always be nonnegative. But for $K = 1$, say, which ensures that the closed-loop system will be asymptotically stable, the above quantity will be nonnegative for $\omega < \sqrt{10.5}$. Consequently, if interest centers around performance of the system only up to a frequency of $\sqrt{10.5}$ rad/sec., we achieve a reduction in sensitivity to plant parameter variation, i.e. if one of the plant poles were to vary, or the plant gain to vary, the presence of the feedback would reduce the effect of this variation.

Above 10.5 rad/sec., we would however expect to see a heightened rather than reduced sensitivity to plant parameter variations.

Problem 7.1-1. With reference to Fig. 7.1-1, compare the output sensitivity to variation in the β block of the closed-loop system, and the $1/(1 + A\beta)$ block of the open-loop system; then, for Fig. 7.1-2, compare the output sensitivity to variation in the $F(s)$ block of the closed-loop system, and the $C(s)$ block of the open-loop system.

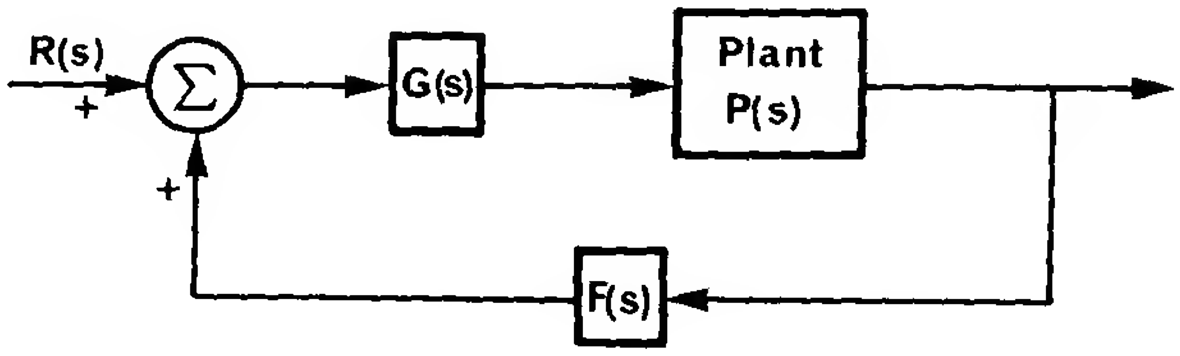


Fig. 7.1-3 Combined series compensator and feedback compensator.

Problem 7.1-2. Consider the arrangement of Fig. 7.1-3, where a series compensator and a feedback control are used. What is the criterion for this system to offer a sensitivity improvement (for plant parameter variations) over an equivalent open-loop system?

Problem 7.1-3. Consider Fig. 7.1-3, with the time-invariant plant and controller replaced by a linear time-varying plant and controller, of impulse responses $p(t, \tau)$ and $f(t, \tau)$, respectively. The operator $\delta(t - \tau) + \int_{-\infty}^{t} p(t, \lambda)f(\lambda, \tau)\, d\lambda$ is known as a causal Volterra operator, and can frequently be shown to possess a causal inverse, $s(t, \tau)$. Show that

$$y_c(t; \mu) - y_c(t; \mu_{\text{nom}}) = \int_{-\infty}^{t} s(t, \tau; \mu)[y_o(\tau; \mu) - y_o(\tau; \mu_{\text{nom}})]\, d\tau.$$

Problem 7.1-4. Consider the arrangements of Fig. 7.1-4, which are intended to be the same as those of Fig. 7.1-2, except for the introduction of a disturbance of Laplace transform $D(s)$ at the plant output. Show that the integral square error in the output due to $D(s)$ for the closed-loop scheme is less than or equal to the same quantity associated with the open-loop scheme if the closed-loop scheme offers sensitivity improvement over the open-loop scheme of the sort discussed.

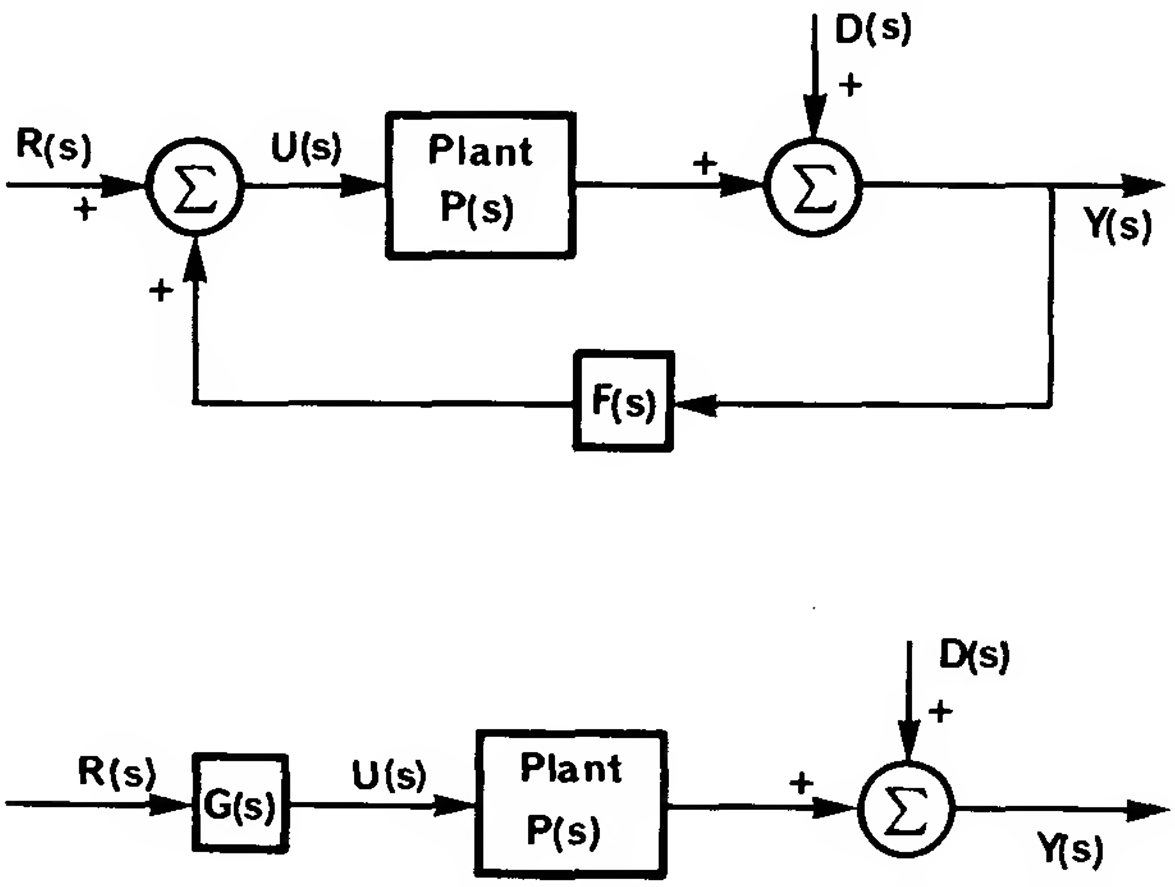


Fig. 7.1-4 Block diagrams showing the introduction of disturbances.

7.2 SENSITIVITY REDUCTION IN OPTIMAL REGULATORS

In this section, we shall apply the results of the preceding section to optimal regulators. For the moment, we shall regard the optimal regulator as consisting of a "plant," with transfer function $-k'_\alpha(sI - F)^{-1}g$, and unity negative feedback. The output is taken to be $-k'_\alpha x$, where x is the state of the system

$$\dot{x} = Fx + gu$$

to which optimal control is applied. The situation is thus that of Fig. 7.2-1.

With these definitions, the criterion for this closed-loop scheme to have a lower sensitivity to variations in F, g, k_α or the plant initial condition than

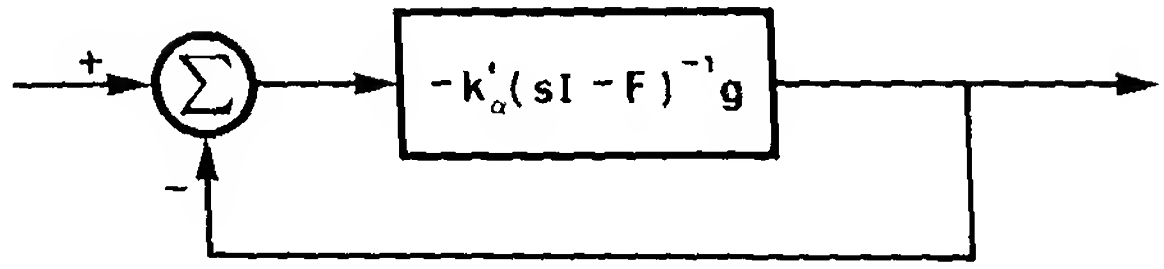


Fig. 7.2-1 Regulator arranged so as to have unity feedback.

an equivalent open-loop system (obtained by cascading a suitable controller with the plant) is the following inequality on the return difference, derived in the last section:

$$|1 - k_\alpha'(j\omega I - F)^{-1}g| \geq 1 \tag{7.2-1}$$

for all real ω.

The fact that the regulator is optimal ensures that this inequality should hold, which is immediate from the fundamental Eq. (5.2-12) of Chapter 5. Thus, we conclude that, with the appropriate definitions of plant and output, optimality implies improvement in sensitivity over an equivalent open-loop system.

We shall now derive a corresponding result for multiple-input systems. We recall the fundamental equation (5.2-7) of Chapter 5:

$$[I - R^{1/2}K_0'(j\omega I - F)^{-1}GR^{-1/2}]^{*\prime}[I - R^{1/2}K_0'(j\omega I - F)^{-1}GR^{-1/2}] \geq I. \tag{7.2-2}$$

With K_0 replaced by K_α, the inequality is also valid. We may regard (7.2-2) (with K_α replacing K_0) as stating that

$$T^{*\prime}(j\omega)T(j\omega) - I \geq 0 \tag{7.2-3}$$

where

$$T(j\omega) = I - R^{1/2}K_\alpha'(j\omega I - F)^{-1}GR^{-1/2}. \tag{7.2-4}$$

We now examine the implications of this inequality using Fig. 7.2-2. The three diagrams in this figure all show equivalent systems, the first showing how the transfer function matrix of the system is achieved by unity negative feedback, the second linking the first and the third, and the third showing how the transfer function matrix is achieved using memoryless transformations of both the input and the output, with a unity negative feedback arrangement linking the transformed input and output. Figure 7.2-3 shows this unity negative feedback part in isolation.

The return difference matrix for the arrangement of Fig. 7.2-3 is precisely the preceding matrix $T(j\omega)$. Consequently, the feedback arrangement of Fig. 7.2-3 has lower sensitivity to plant parameter variations than an equivalent open-loop system.

There is also a simple consequence following from the equivalences of Fig. 7.2-2, which is relevant for the first arrangement shown in this figure. We retain the notation of the previous section and introduce the new variable

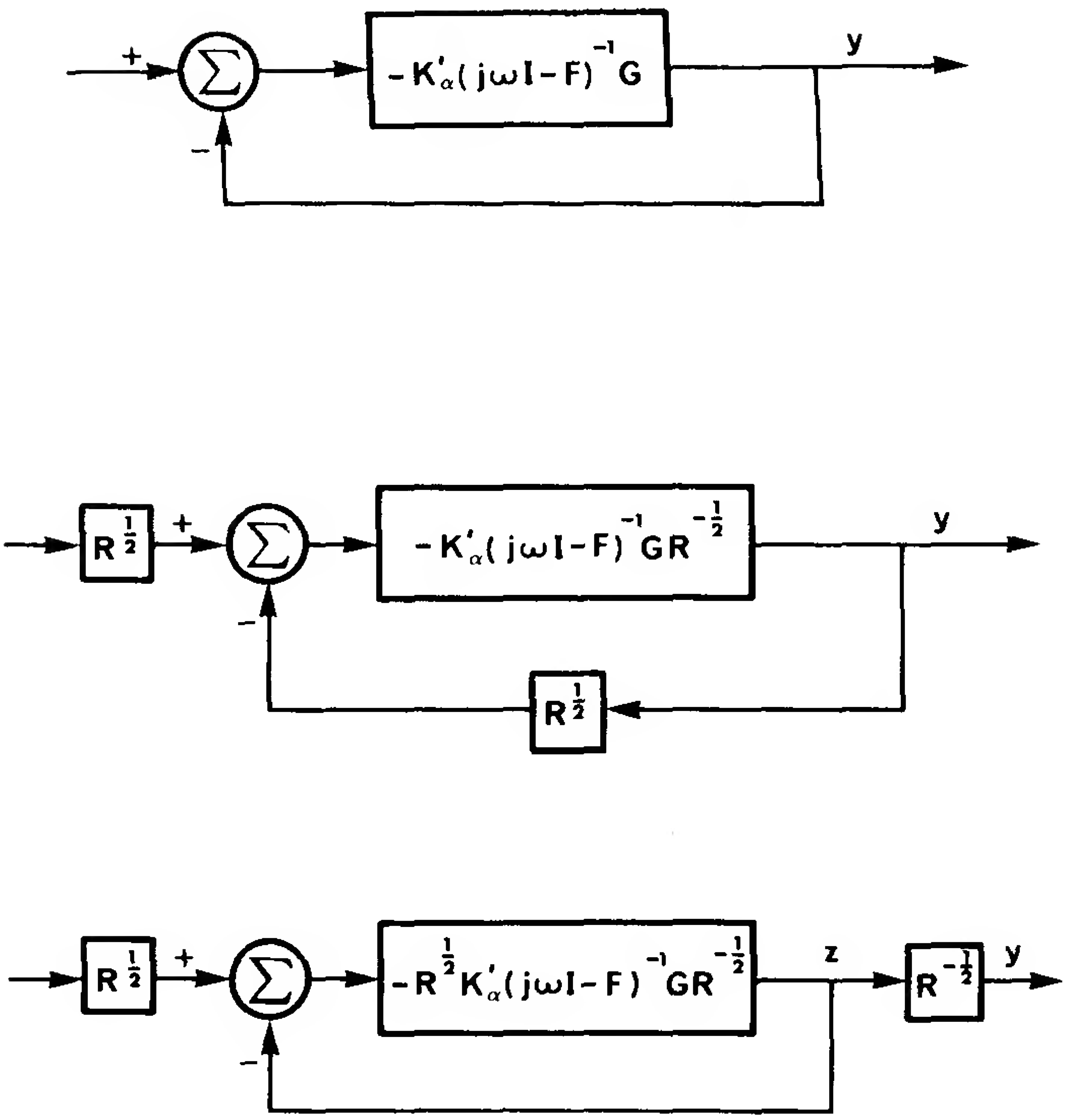


Fig. 7.2-2 Equivalent systems.

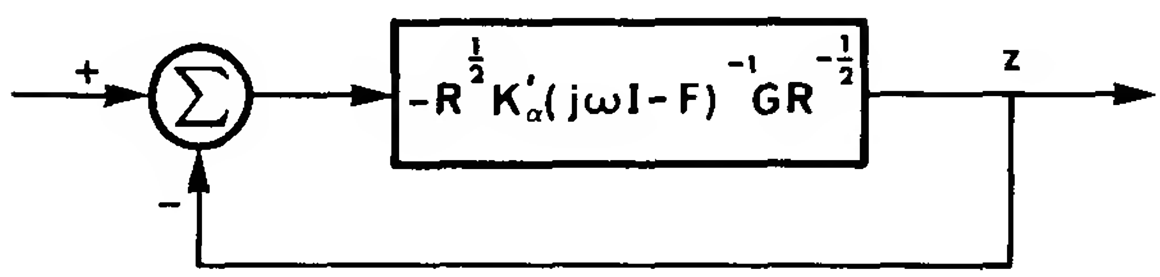


Fig. 7.2-3 Unity feedback system with output $z = R^{-1/2}y$.

z, defined by $z = R^{-1/2}y$. The sensitivity reduction associated with the arrangement of Fig. 7.2-3 implies that

$$\int_{t_0}^{t_1} [z_c(\mu) - z_c(\mu_{\text{nom}})]'[z_c(\mu) - z_c(\mu_{\text{nom}})]\, dt$$
$$\leq \int_{t_0}^{t_1} [z_o(\mu) - z_o(\mu_{\text{nom}})]'[z_o(\mu) - z_o(\mu_{\text{nom}})]\, dt$$

for all t_0, t, and variations in z_o. Since $y = R^{-1/2}z$, it follows that

$$\int_{t_0}^{t_1} [y_c(\mu) - y_c(\mu_{\text{nom}})]'R[y_c(\mu) - y_c(\mu_{\text{nom}})]\, dt$$
$$\leq \int_{t_0}^{t_1} [y_o(\mu) - y_o(\mu_{\text{nom}})]'R[y_o(\mu) - y_o(\mu_{\text{nom}})]\, dt.$$

This is certainly meaningful. Suppose we merely replaced the original definition of reduction of sensitivity, viz.,

$$\int_{t_0}^{t_1} [y_c(\mu) - y_c(\mu_{\text{nom}})]'[y_c(\mu) - y_c(\mu_{\text{nom}})]\,dt$$
$$\leq \int_{t_0}^{t_1} [y_o(\mu) - y_o(\mu_{\text{nom}})]'[y_o(\mu) - y_o(\mu_{\text{nom}})]\,dt$$

for all t_0, t_1 and variations in y_o, by one where we can weight the various entries of the error vectors separately. Namely,

$$\int_{t_0}^{t_1} [y_c(\mu) - y_c(\mu_{\text{nom}})]'W[y_c(\mu) - y_c(\mu_{\text{nom}})]\,dt$$
$$\leq \int_{t_0}^{t_1} [y_o(\mu) - y_o(\mu_{\text{nom}})]'W[y_o(\mu) - y_o(\mu_{\text{nom}})]\,dt$$

for some positive definite W, all t_0, t_1 and variations in y_o. We may then conclude that the first scheme of Fig. 7.2-2 offers a sensitivity reduction over an equivalent open-loop scheme—i.e., with appropriate definition of plant and output, the multiple input regulator is a feedback scheme where plant parameter sensitivity is reduced.

At this stage, we might examine the usefulness of what we have established. Figure 7.2-4 shows how we may expect to use an optimal regulator. The true output of the regulator is not—at least, normally—$-K'_\alpha x$, but rather $H'x$. Now, the above remarks have established that variations in F, G and K_α give rise to reduced sensitivity in $K'_\alpha x$ rather than $H'x$; consequently, we do not have a direct result concerning reduction in sensitivity of the regulator output.

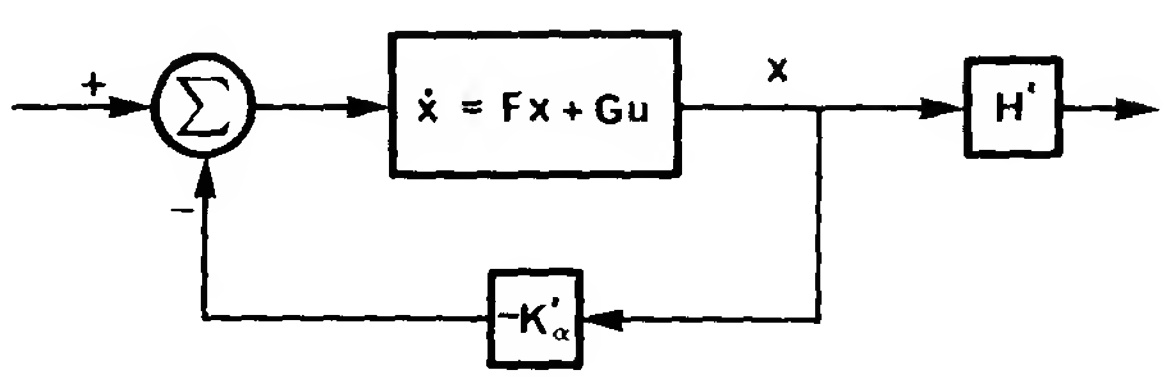


Fig. 7.2-4 Regulator diagram.

To be sure, it is useful to know that the sensitivity of $K'_\alpha x$ is reduced, in the sense that (7.2-5), rewritten as follows,

$$\int_{t_0}^{t_1} [x_c(\mu) - x_c(\mu_{\text{nom}})]'K_\alpha R K'_\alpha[x_c(\mu) - x_c(\mu_{\text{nom}})]\,dt$$
$$\leq \int_{t_0}^{t_1} [x_o(\mu) - x_o(\mu_{\text{nom}})]'K_\alpha R K'_\alpha[x_o(\mu) - x_o(\mu_{\text{nom}})]\,dt, \qquad (7.2\text{-}6)$$

should hold for all t_0, t_1 and variations in x_o. If, in fact, $K'_\alpha R K'_\alpha$ were positive definite (requiring the number of inputs to at least equal the number of states),

(7.2-6) would actually imply that the states themselves have the sensitivity reduction property. But, in general, $K_\alpha R K'_\alpha$ is merely nonnegative definite, and such a statement will not normally be true. Nor, of course, will (7.2-6) imply a relation like

$$\int_{t_0}^{t_1} [x_c(\mu) - x_c(\mu_{\text{nom}})]' HWH' [x_c(\mu) - x_c(\mu_{\text{nom}})]\, dt$$
$$\leq \int_{t_0}^{t_1} [x_o(\mu) - x_o(\mu_{\text{nom}})]' HWH' [x_o(\mu) - x_o(\mu_{\text{nom}})]\, dt$$

for some positive definite W, all t_o, t_1, and variations in x_o.

There is, however, one sensitivity result involving the states x directly (rather than $K'_\alpha x$), which applies for single-input systems, when there are restrictions placed on the type of parameter variations. We shall now derive this result.

We consider the arrangement of Fig. 7.2-5 with u restricted to being a scalar; for the purposes of applying the sensitivity analysis of the last section, we note that the "plant" has transfer function matrix $(sI - F)^{-1}g$. The quantities we called $Y_o(s; \mu_{\text{nom}})$, $Y_c(s; \mu)$, etc., in the last section will be denoted here by $X_o(s; \mu_{\text{nom}})$, $X_c(s; \mu)$, etc., since in this case the plant output vector is actually a state vector.

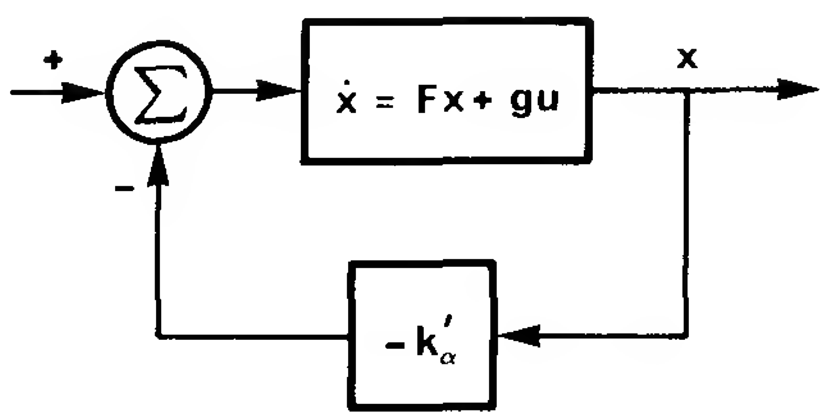


Fig. 7.2-5 Plant with state vector as output.

The parameter variations allowed are those where the matrix F is replaced by $F + gl'$ for some vector l, all of whose entries are small. We shall show that with this restricted parameter variation, the sensitivity of the states is improved by the optimal feedback.

Before doing so, we indicate other types of parameter variation equivalent to that stated. (Thus, the original restriction need not appear quite so narrow.) First, instead of changing F to $F + gl'$, we may change $-k'_\alpha$ to $-k'_\alpha - m'$, where the entries of the vector m are small. This parameter variation is equivalent to one where F is changed to $F + gm'$ and $-k'_\alpha$ is left unaltered (see Fig. 7.2-6, which illustrates the equivalence). Second, instead of changing F to $F + gl'$, we may change $-k'_\alpha$ to $-(1 + \gamma)k'_\alpha$, where γ is a small number. In so doing, we are altering the whole loop gain slightly, by the same factor for all frequencies. This is clearly a special case of the preceding first equivalence, since $-(1 + \gamma)k'_\alpha = -k'_\alpha - m'$ with $m = \gamma k_\alpha$. With γ small, the entries of m are small. Therefore, this sort of parameter

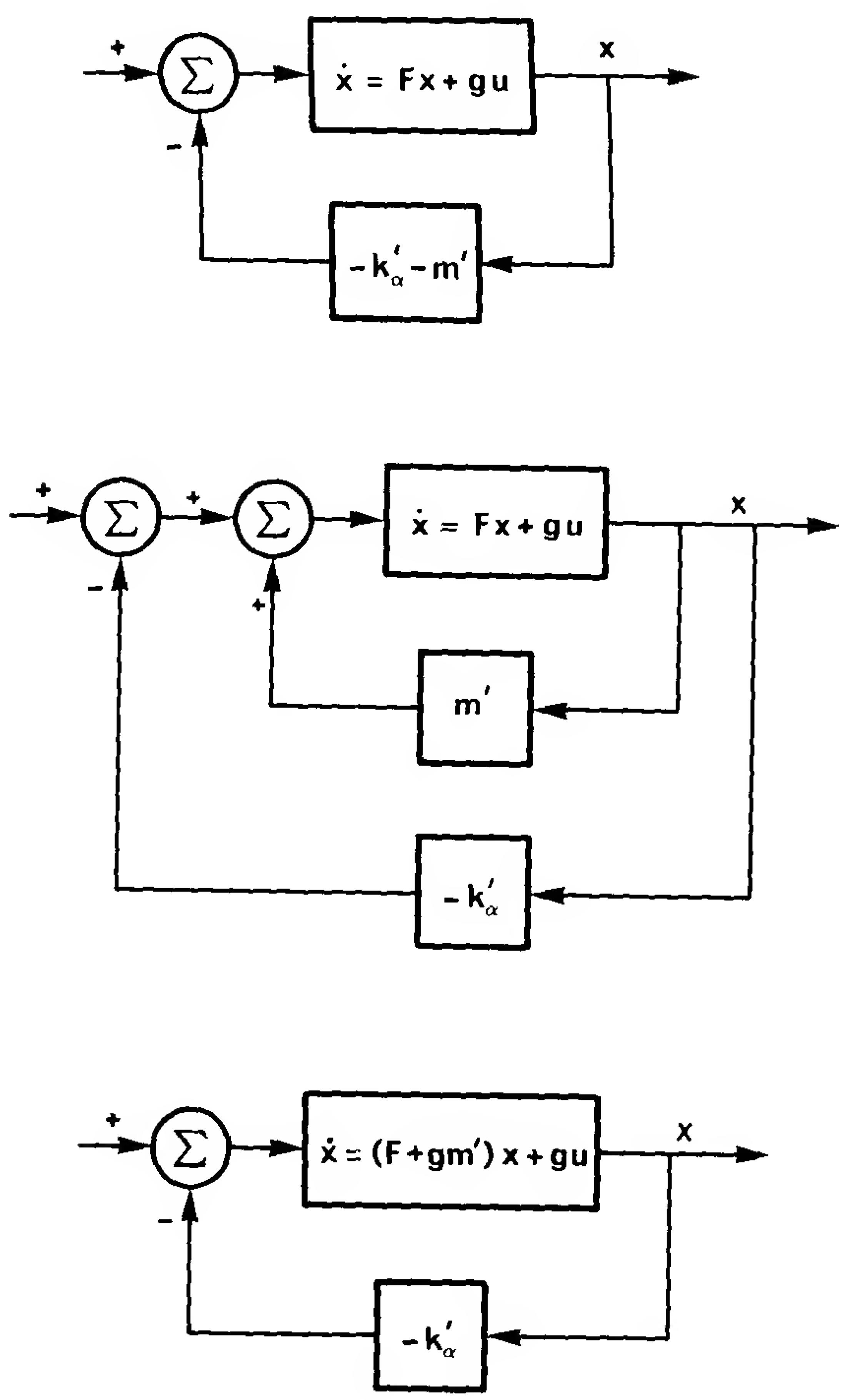


Fig. 7.2-6 Equivalence of certain parameter variations.

variation is also within the ambit of that originally allowed. It is also easy to argue that a variation in g, from g to $(1 + \gamma)g$ where γ is very small, is equivalent to a parameter variation of the original type.

We now show that, with the sorts of parameter variation allowed, we obtain the desired sensitivity reduction. We recall first Eqs. (7.1-2) and (7.1-3) of the last section in the original notation, and repeat them for convenience:

$$Y_o(\mu) - Y_0(\mu_{\text{nom}}) = [P(\mu) - P(\mu_{\text{nom}})]U_o(\mu_{\text{nom}}) + Z(\mu) - Z(\mu_{\text{nom}})$$

and

$$\begin{aligned}[I + P(\mu)F][Y(\mu) - Y(\mu_{\text{nom}})]\\ = [P(\mu) - P(\mu_{\text{nom}})]U_c(\mu_{\text{nom}}) + Z(\mu) - Z(\mu_{\text{nom}}),\end{aligned}$$

with, also, $U_o(\mu_{\text{nom}}) = U_c(\mu_{\text{nom}})$.

Making the various applicable specializations, we get

$$X_o(s;\mu) - X_o(s;\mu_{\text{nom}}) = [(sI - F - gl')^{-1}g - (sI - F)^{-1}g]U_o(s;\mu_{\text{nom}}) \tag{7.2-7}$$

and

$$\begin{aligned}[I - (sI - F - gl')^{-1}gk'_\alpha][X_c(s;\mu) - X_c(s;\mu_{\text{nom}})]\\ = [(sI - F - gl')^{-1}g - (sI - F)^{-1}g]U_o(s;\mu_{\text{nom}}).\end{aligned} \tag{7.2-8}$$

We shall later establish the two identities

$$(sI - F - gl')^{-1}g - (sI - F)^{-1}g = [l'(sI - F)^{-1}g](sI - F - gl')^{-1}g \tag{7.2-9}$$

and

$$\begin{aligned}[I - (sI - F - gl')^{-1}gk'_\alpha]^{-1}(sI - F - gl')^{-1}g\\ = [1 - k'_\alpha(sI - F - gl')^{-1}g]^{-1}(sI - F - gl')^{-1}g.\end{aligned} \tag{7.2-10}$$

Inserting (7.2-9) into (7.2-7) and (7.2-8), we obtain

$$X_o(s;\mu) - X_o(s;\mu_{\text{nom}}) = [l'(sI - F)^{-1}g](sI - F - gl')^{-1}gU_o(s;\mu_{\text{nom}}) \tag{7.2-11}$$

and

$$\begin{aligned}X_c(s;\mu) - X_c(s;\mu_{\text{nom}}) = [l'(sI - F)^{-1}g][I - (sI - F - gl')^{-1}gk'_\alpha]^{-1}\\ \times (sI - F - gl')^{-1}U_o(s;\mu_{\text{nom}})\end{aligned} \tag{7.2-12}$$

and inserting (7.2-10) into (7.2-12), we have

$$\begin{aligned}&X_c(s;\mu) - X_c(s;\mu_{\text{nom}})\\ &\quad= [1 - k'_\alpha(sI - F - gl')^{-1}g]^{-1}[l'(sI - F)^{-1}g](sI - F - gl')^{-1}gU_o(s;\mu_{\text{nom}})\\ &\quad= [1 - k'_\alpha(sI - F - gl')^{-1}g]^{-1}[X_o(s;\mu) - X_o(s;\mu_{\text{nom}})]\end{aligned}$$

on using (7.2-11).

So far, we have not insisted that the entries of l be small; all equations have been exact. But now, by requiring that the entries of l be small, we can make the same sort of approximation as in the last section to conclude that (approximately),

$$X_c(s;\mu) - X_c(s;\mu_{\text{nom}}) = [1 - k'_\alpha(sI - F)^{-1}g]^{-1}[X_o(s;\mu) - X_o(s;\mu_{\text{nom}})]. \tag{7.2-13}$$

We recognize $1 - k'_\alpha(sI - F)^{-1}g$ again to be the return difference. Thus, (7.2-13) guarantees that for each entry of the state vector (and thus for the whole state vector) the closed-loop system provides a lower sensitivity to parameter variation of the sort specified than an equivalent open-loop system.

Before commenting further, we establish the two matrix identities used previously. First, we deal with (7.2-9):

$$\begin{aligned}(sI - F - gl')^{-1} &- (sI - F)^{-1} \\ &= (sI - F - gl')^{-1}[(sI - F) - (sI - F - gl')](sI - F)^{-1} \\ &= (sI - F - gl')^{-1}gl'(sI - F)^{-1}.\end{aligned}$$

Equation (7.2-9) follows by multiplying in the right by g. Second, for (7.2-10), we have

$$\begin{aligned}[I - (sI &- F - gl')^{-1}gk'_\alpha](sI - F - gl')^{-1}g \\ &= (sI - F - gl')^{-1}g - (sI - F - gl')^{-1}g[k'_\alpha(sI - F - gl')^{-1}g] \\ &= [1 - k'_\alpha(sI - F - gl')^{-1}g](sI - F - gl')^{-1}g.\end{aligned}$$

Equation (7.2-10) follows immediately.

Because of the fact that each entry of the state vector undergoes the sensitivity reduction we have discussed, it follows that if $y = h'x$ is the actual output of the regulator with feedback, the sensitivity reduction property of the closed-loop scheme extends to y. Quantitatively, multiplication of (7.2-13) on the left by h' yields

$$Y_c(s; \mu) - Y_c(s; \mu_{\text{nom}}) = [1 - k'_\alpha(sI - F)^{-1}g]^{-1}[Y_o(s; \mu) - Y_o(s; \mu_{\text{nom}})]. \tag{7.2-14}$$

We stress that y is of more general form here than in, say, Eq. (7.2-5), where we had $y = -k'_\alpha x$. We remark again that (7.2-14) is only valid for single-input systems, with the restricted type of parameter variations discussed previously.

We round off the theory by discussing the degree of sensitivity improvement obtained by varying the parameter α in the regulator design, with the F, G, Q, and R matrices remaining invariant; each α gives rise to a control law K_α, and we wish to show that different values of α will give rise to different amounts of sensitivity improvement, in the sense discussed at the end of the previous section. To see this, we recall first the fundamental relation (5.2-6) of Chapter 5, repeated here for convenience:

$$\begin{aligned}[I - R^{-1/2}G'(-j\omega I &- F')^{-1}K_0R^{1/2}][I - R^{1/2}K'_0(j\omega I - F)^{-1}GR^{-1/2}] \\ &= I + R^{-1/2}G'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}GR^{-1/2}.\end{aligned} \tag{7.2-15}$$

The corresponding equation when K_0 is replaced by K_α, obtained as described in Chapter 5, Sec. 5.2 is

$$\begin{aligned}[I - R^{-1/2}G'(-j\omega I &- F')^{-1}K_\alpha R^{1/2}][I - R^{1/2}K'_\alpha(j\omega I - F)^{-1}GR^{-1/2}] \\ &= I + R^{-1/2}G'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}GR^{-1/2} \\ &\quad + R^{-1/2}G'(-j\omega I - F')^{-1}2P_\alpha(j\omega I - F)^{-1}GR^{-1/2}.\end{aligned} \tag{7.2-16}$$

With the definitions

$$T_1(j\omega) = I - R^{1/2}K'_0(j\omega I - F)^{-1}GR^{-1/2}$$

and

$$T_2(j\omega) = I - R^{1/2}K'_\alpha(j\omega I - F)^{-1}GR^{-1/2},$$

it is immediate from (7.2-15) and (7.2-16) that

$$T_2^{*\prime}(j\omega)T_2(j\omega) \geq T_1^{*\prime}(j\omega)T_1(j\omega). \tag{7.2-17}$$

Therefore, with interpretations of T_1 and T_2 as matrix return differences in the various situations just discussed, we conclude, following the basic theory of the last section, that a nonzero α offers a greater degree of sensitivity improvement than a zero α.

An interesting example appearing in [5] applies some of these ideas to a pitch-axis control system for a high-performance aircraft. The plant equations are third order and linear, but contain parameter variations arising from different dynamic pressures. With

x_1 = Angle of attack,

x_2 = Rate of pitch angle,

x_3 = Incremental elevator angle, and

u = Control input into the elevator actuator,

the state equations are

$$\dot{x} = \begin{bmatrix} -0.074 & 1 & -0.012 \\ -8.0 & -0.055 & -6.2 \\ 0 & 0 & -6.6667 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 6.6667 \end{bmatrix} u$$

for a dynamic pressure of 500, whereas for a dynamic pressure of 22.1, they become

$$\dot{x} = \begin{bmatrix} -0.0016 & 1.0 & -0.0002 \\ -0.1569 & -0.0015 & -0.1131 \\ 0 & 0 & -6.6667 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 6.6667 \end{bmatrix} u.$$

The performance index minimized is $\int_{t_0}^{\infty} (u^2 + x'Qx)\,dt$ for two different matrices Q—viz.,

$$Q_1 = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad Q_2 = \begin{bmatrix} 2 \times 10^6 & 0 & 0 \\ 0 & 4 \times 10^5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Two optimal control laws $u = k'_1x$ and $u = k'_2x$ are calculated, based on the first state equation in both cases, and on performance indices including Q_1 and Q_2, respectively. The control laws are then used not just for the first state equation (for which they are optimal) but for the second state equation, for which they are not optimal. The following table summarizes the results achieved; the sensitivity improvement refers to the sensitivity in x_1.

State Equation	Control Law	Step Response	Sensitivity Improvement
First	k_1	Very Good	Moderate
Second	k_1	Very Poor	
First	k_2	Good	Huge
Second	k_2	Acceptable	

Although the theory presented suggests that in this case the feedback optimal control is the only variable for which sensitivity improvement occurs, we see here that the sensitivity improvment in the control is evidently linked with sensitivity improvement in the response of x_1. This example also illustrates the notion that large Q lead to less sensitive closed-loop systems than smaller Q. Although the entries of the F matrix undergo substantial change, the response of x_1 to a typical initial condition varies but little, when the design based on Q_2 is used.

The first sensitivity result for linear optimal systems was achieved in [6], where it was noted that the quantity $-k'x$ occurring in the single-input problem could be associated with a sensitivity reduction result. The corresponding result for multiple-input systems is derived in [7]. Versions of these results for time-varying optimal regulators are described in [8].

The latter material of this section, dealing with the state sensitivity question, appears in [5], where the necessity for constraining the allowed sorts of parameter variations is not directly indicated, and also in [9]. In both references, a restriction is made on the F matrix and g vector of the system, which we have avoided here.

An entirely different sensitivity question arises when one asks what is the sensitivity of the optimal performance index of an optimal system—in particular, the linear regulator—to parameter variation, and how do the sensitivities for open- and closed-loop implementations of the optimal control vary. This question is discussed in [9]. The somewhat surprising result for the linear regulator is that the two sensitivities are the same. Various other sensitivity problems are discussed in reference [10].

Problem 7.2-1. Suppose an optimal control law $u = k_\alpha' x$ is found for a single-input system with F and g given by

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & & & 1 \\ -a_1 & -a_2 & -a_3 & \cdot & \cdot & -a_n \end{bmatrix} \qquad g = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix}.$$

Show that the sensitivity of the states of the system is less for closed- than for open-loop control implementations, provided that only the two following types of parameter variation are permitted:

1. One or more of the a_i are varied slightly.
2. The nonzero entry of g is varied slightly.

Problem 7.2-2. Attempt to generalize the development of Eqs. (7.2-7) through (7.2-13) to multiple-input systems, and show where the generalization falls down.

Problem 7.2-3. Suppose two optimal regulator designs are carried out with the same F, G, Q, and R matrices and with two values of α. Call them α_1, α_2, with the constraint $\alpha_1 > \alpha_2 > 0$. Show that the α_1 design offers greater sensitivity improvement than the α_2 design. [*Hint:* Recall that $P_{\alpha_1} - P_{\alpha_2}$ is positive definite at a critical point in your calculations.]

7.3 THE INVERSE REGULATOR PROBLEM

In this section, we consider a problem of somewhat theoretical interest, the solution of which nevertheless sheds light on the preceding sensitivity ideas. In the next section we shall apply the solution of the problem in a further discussion of the "pole-positioning problem," i.e., the task of choosing an optimal feedback law that also achieves desired pole positions.

We ask the following question. Given a linear system

$$\dot{x} = Fx + Gu \tag{7.3-1}$$

with control law

$$u = K'x, \tag{7.3-2}$$

is there a positive definite R and nonnegative definite Q such that (7.3-2) is the optimal control law for the system (7.3-1) with performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u'Ru + x'Qx)\,dt? \tag{7.3-3}$$

Obviously, for arbitrary F, G, and K we could not expect (7.3-2) to be optimal for some Q and R. Accordingly, we are led to ask whether certain constraints can be imposed on F, G, and K, which will ensure that (7.3-2) is an optimal control; furthermore, we might be interested in computing the Q and R matrices appearing in (7.3-3).

We shall give a detailed consideration of this problem only for the scalar input case, because the manipulations required in the vector input case are particularly intricate. We shall, however, subsequently state a result for the vector input case.

More or less, the only regulator systems of interest are those that are

completely controllable and asymptotically stable. Accordingly, we shall suppose the following.

ASSUMPTION 7.3-1. The pair $[F, g]$ is completely controllable.

ASSUMPTION 7.3-2. The eigenvalues of $F + gk'$ all have negative real parts. [Recall that the closed-loop system is $\dot{x} = (F + gk')x$.]

It turns out that to keep the calculations simple, it is convenient to impose one further assumption temporarily.

ASSUMPTION 7.3-3. The pair $[F, k]$ is completely observable. (Thus, the feedback control function can only be identically zero if the system state is identically zero.)

Problem 7.3-1 asks for a demonstration that this constraint may be removed.

We now claim that the scheme (7.3-1), (7.3-2), with scalar input, is optimal [i.e., there exist a Q and an R with appropriate constraints such that (7.3-2) is the optimal control for the performance index (7.3-3)], provided that,

$$|1 - k'(j\omega I - F)^{-1}g| \geq 1 \tag{7.3-4}$$

for all real ω. Equation (7.3-4) is precisely the sensitivity improvement condition for the closed-loop system defined by (7.3-1) and (7.3-2). Moreover, if (7.3-1) and (7.3-2) are optimal, then (7.3-4) is known to hold, by the material of Sec. 7.2. Consequently, with the assumptions of complete controllability and asymptotic stability just given, (7.3-4) is both necessary and sufficient. This result constitutes a connection between the modern approach to linear control system design and an approach of classical control of very long standing, to which we referred at the start of the chapter. Although the full implications of (7.3-4) in respect to sensitivity reduction have not been known until recently, a sufficient number of them have been known for many years for designers to have attempted to ensure the satisfaction of (7.3-4) or some suitable variant using classical procedures.

Equation (7.3-4) is a condition on the transfer function $k'(j\omega I - F)^{-1}g$. Accordingly, it is invariant with respect to changes of the coordinate basis used for defining the state vector in (7.3-1) and (7.3-2)—i.e., if F is replaced by $F_1 = TFT^{-1}$, g by $g_1 = Tg$, and k' by $k'_1 = k'T^{-1}$, we have that $k'_1(j\omega I - F_1)^{-1}g_1 = k'(j\omega I - F)^{-1}g$. Therefore, there is no loss of generality in assuming a special coordinate basis; we shall avail ourselves of this opportunity and require (as we may by the complete controllability constraint, see Appendix B) that the coordinate basis be chosen to yield

$$F = \begin{bmatrix} 0 & 1 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \\ -a_1 & -a_2 & -a_3 & \cdot & \cdot & -a_n \end{bmatrix} \qquad g = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}. \tag{7.3-5}$$

We assume that

$$k = [k_1 \quad k_2 \quad \cdots \quad k_n]. \tag{7.3-6}$$

Make the definition

$$\det(sI - F) = \psi(s). \tag{7.3-7}$$

Then the following three relations hold:

$$\psi(s) = s^n + a_n s^{n-1} + \cdots + a_1, \tag{7.3-8}$$

$$(sI - F)^{-1}g = \frac{1}{\psi(s)} \begin{bmatrix} 1 \\ s \\ s^2 \\ \cdot \\ \cdot \\ s^{n-1} \end{bmatrix},$$

and

$$k'(sI - F)^{-1}g = \frac{k_n s^{n-1} + k_{n-1}s^{n-2} + \cdots + k_1}{\psi(s)}. \tag{7.3-9}$$

We shall now use the relation (7.3-4) to construct a Q matrix; we shall take the R matrix in the performance index (7.3-3) to be unity.

With the further definition

$$p(s) = s^n + (a_n - k_n)s^{n-1} + \cdots + (a_1 - k_1), \tag{7.3-10}$$

it follows from (7.3-4) that

$$[1 - k'(-j\omega I - F)^{-1}g][1 - k'(j\omega I - F)^{-1}g] = \frac{p(-j\omega)p(j\omega)}{\psi(-j\omega)\psi(j\omega)} \geq 1, \tag{7.3-11}$$

or

$$\frac{p(-j\omega)p(j\omega) - \psi(-j\omega)\psi(j\omega)}{\psi(-j\omega)\psi(j\omega)} \geq 0. \tag{7.3-12}$$

Consequently, the numerator polynomial of this expression can be factored as

$$p(-j\omega)p(j\omega) - \psi(-j\omega)\psi(j\omega) = m(-j\omega)m(j\omega) \tag{7.3-13}$$

for some polynomial

$$m(s) = d_n s^{n-1} + d_{n-1} s^{n-2} + \cdots + d_1 \tag{7.3-14}$$

with all zeros possessing negative real parts. [Note that $m(s)$ has degree at most $n - 1$, as inspection of the coefficients of the s^n terms of $\psi(s)$ and $p(s)$ will show.]

Equations (7.3-11) through (7.3-13) combine to yield

$$\begin{aligned}[1 - k'(-j\omega I - F)^{-1}g]&[1 - k'(j\omega I - F)^{-1}g] \\ &= 1 + \frac{m(-j\omega)m(j\omega)}{\psi(-j\omega)\psi(j\omega)} \\ &= 1 + g'(-j\omega I - F)^{-1}dd'(j\omega I - F)^{-1}g \end{aligned} \tag{7.3-15}$$

where the vector d is given by

$$d' = [d_1 \quad d_2 \quad \cdots \quad d_n] \tag{7.3-16}$$

and we have used the relation, analogous to (7.3-9), that

$$d'(j\omega I - F)^{-1}g = \frac{1}{\psi(j\omega)}[d_n(j\omega)^{n-1} + \cdots + d_1] = \frac{m(j\omega)}{\psi(j\omega)}.$$

Now, if k were the optimal control law corresponding to $R = I$ and a certain matrix Q in (7.3-3), the results of Sec. 5.2 [see the fundamental equation (5.2-9)], would require that

$$\begin{aligned}[1 - k'(-j\omega I - F)^{-1}g]&[1 - k'(j\omega I - F)^{-1}g] \\ &= 1 + g'(-j\omega I - F')^{-1}Q(j\omega I - F)^{-1}g.\end{aligned} \tag{7.3-17}$$

The similarity between (7.3-15) and (7.3-17) then suggests that with $Q = dd'$ in (7.3-3), the control law k *might* be optimal. *Note that mere satisfaction of* (7.3-15) *by some* k *will not necessarily guarantee the optimality of* k*;* Eq. (7.3-15) is a necessary, but not a sufficient, condition on k. Nevertheless, we can now prove that k is, in fact, optimal for the choice $Q = dd'$, $R = 1$.

Suppose k is not optimal, and that the optimal control is, instead, $u = \bar{k}'x$. Using the fundamental result of Chapter 5, Eq. (5.2-9), it follows that

$$\begin{aligned}[1 - \bar{k}'(-j\omega I - F)^{-1}g]&[1 - \bar{k}'(j\omega I - F)^{-1}g] \\ &= 1 + g'(-j\omega I - F')^{-1}dd'(j\omega I - F)^{-1}g.\end{aligned} \tag{7.3-18}$$

Comparison of (7.3-15) and (7.3-18) gives immediately that

$$\begin{aligned}[1 - \bar{k}'(-j\omega I - F)^{-1}g]&[1 - \bar{k}'(j\omega I - F)^{-1}g] \\ &= [1 - k'(-j\omega I - F)^{-1}g][1 - k'(j\omega I - F)^{-1}g].\end{aligned} \tag{7.3-19}$$

We shall now show that this equation implies that $k = \bar{k}$.

Define a polynomial $\bar{p}(s)$ in the same way as $p(s)$ [see (7.3-10)] except that k_i in the definition is replaced by $\bar{k}_i$ for each i. Thus, (7.3-18) and (7.3-19)

imply that

$$\frac{p(-j\omega)p(j\omega)}{\psi(-j\omega)\psi(j\omega)} = \frac{\bar{p}(-j\omega)\bar{p}(j\omega)}{\psi(-j\omega)\psi(j\omega)} = 1 + \frac{m(-j\omega)m(j\omega)}{\psi(-j\omega)\psi(j\omega)}. \tag{7.3-20}$$

Since the coefficients of $p(s)$ and $\bar{p}(s)$ are the entries of the vectors k and $\bar{k}$, the proof will be complete if we can show that $p(s) = \bar{p}(s)$. Now it is easy to check that

$$p(s) = \det(sI - F - gk') \tag{7.3-21}$$

and

$$\bar{p}(s) = \det(sI - F - g\bar{k}'). \tag{7.3-22}$$

[This follows from the definition of the polynomials, together with a minor modification of the calculations yielding (7.3-8) from (7.3-5) and (7.3-7).] Because $F + gk'$ is known to have all eigenvalues with negative real parts, $p(s)$ is uniquely determined as the spectral factor of the polynomial $p(-j\omega)p(j\omega)$, which has all zeros in the left half-plane.

If we can show that $\bar{p}(s)$ has the property that it, too, has all zeros in the left half-plane, it will follow from (7.3-20) that $p(s) = \bar{p}(s)$. From (7.3-22), it is clear that $\bar{p}(s)$ will have all its zeros in the left half-plane if the optimal closed-loop system, $\dot{x} = (F + g\bar{k}')x$, is asymptotically stable. A sufficient condition for this is that the pair $[F, d]$ be completely observable.

Consequently, the desired result follows if the pair $[F, d]$ is proved to be completely observable. We now show by contradiction that this must be the case. Thus, suppose $[F, d]$ is not completely observable. Then the transfer function $d'(sI - F)^{-1}g$, when expressed as a ratio of two polynomials with no common factors, must (see Appendix B) have denominator polynomial of degree less than the dimension of F. Since $d'(sI - F)^{-1}g = m(s)/\psi(s)$, and the degree of ψ is equal to the dimension of F, it follows that m and ψ have a common factor. Because $m(s)$ has all zeros with negative real part, the common factor must possess the same property. Equation (7.3-20) implies that $p(-s)p(s)$ and $\psi(s)$ must have the same common factor as $m(s)$ and $\psi(s)$, and since $p(s)$ has all its zeros with negative real part, $p(s)$ and $\psi(s)$ must have this same common factor. Therefore, $\psi(s) - p(s)$ and $\psi(s)$ have a common factor, and the transfer function

$$k'(sI - F)^{-1}g = \frac{\psi(s) - p(s)}{\psi(s)} \tag{7.3-23}$$

is expressible as a ratio of two polynomials, with the denominator polynomial of lower degree than the dimension of F. This (see Appendix B again) contradicts the complete controllability of $[F, g]$ and complete observability of $[F, k]$. Therefore, our original premise that $[F, d]$ is not completely observable is untenable, and the proof is complete.

Let us now summarize the preceding material. We posed the problem of deciding when the feedback system of (7.3-1) and (7.3-2) resulted from an

optimal regulator design. Then, with several restrictions—viz., complete controllability, asymptotic stability, and a further restriction of complete observability, which Problem 7.3-1 removes—we claimed that the frequency domain condition requiring the return difference to be never less than unity was necessary and sufficient for optimality. Using the frequency domain condition, we constructed a performance index, which, we claimed, led to an optimal control the same as the one prescribed. We were then able to verify this claim. This result was first established in [6].

There exists a corresponding result for multiple-input systems that we can now state. We suppose that we are given the scheme of (7.3-1) and (7.3-2), with complete controllability of $[F, G]$ and asymptotic stability of the closed-loop system. We suppose too that $[F, K]$ is completely observable, although this restriction is presumably removable. Then, if for some positive definite symmetric R we have

$$[I - R^{1/2}K_0'(j\omega I - F)^{-1}GR^{-1/2}]^{*\prime}[I - R^{1/2}K_0'(j\omega I - F)^{-1}GR^{-1/2}] \geq I, \tag{7.3-24}$$

it can be shown that there is a nonnegative definite symmetric Q such that the prescribed control law (7.3-2) is optimal for the performance index (7.3-3), [7].

We remark again that, other than in the application given in the next section, the utility of this section lies in the fact that new illumination of the sensitivity reduction result is provided. In broad terms, feedback systems where sensitivity is reduced are the same as optimally designed systems.

Problem 7.3-1. Consider the system

$$\dot{x} = \begin{bmatrix} F_{11} & 0 \\ F_{12} & F_{22} \end{bmatrix} x + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} u$$

with control law

$$u = [k_1' \quad 0]x$$

where $[F_{11}, g_1]$ is completely controllable, $[F_{11}, k_1]$ is completely observable, and the closed-loop system is asymptotically stable. Suppose, moreover, that the magnitude of the return difference is no smaller than unity at all frequencies. Show that there is a matrix Q such that the control law is optimal for the performance index $\int_{t_0}^{\infty} (u^2 + x'Qx)\,dt$ [*Hint:* Start by considering the feedback system $\dot{x}_1 = F_{11}x_1 + g_1 u$, $u = k_1'x_1$.]

Problem 7.3-2. In the course of establishing the result that

$$|1 - k'(j\omega I - F)^{-1}g| \geq 1$$

for all real ω implies optimality of the control law $u = k'x$ for the system $\dot{x} = Fx + gu$, we constructed a particular performance index for which $u = k'x$ was the minimizing control. In this performance index, the Q matrix was of the form dd' for a vector d. Show the following:

1. In general, the choice of d is not unique, but only a finite number of choices are possible.
2. If Q is not restricted to being of rank 1, there are an infinite number of Q matrices for which $u = k'x$ is the associated optimal control.

Problem 7.3-3. Suppose we are given a system $\dot{x} = Fx + gu$ with feedback law $u = k'x$. Suppose also that the pair $[F, g]$ is completely controllable, that the pair $[F, k]$ is completely observable, and that the closed-loop system $\dot{x} = (F + gk')x$ has degree of stability greater than α. Show that $u = k'x$ is an optimal control law for a performance index of the form $\int_{t_0}^{\infty} e^{2\alpha t}(u^2 + x'dd'x)\,dt$, provided that

$$|1 - k'[(j\omega - \alpha)I - F]^{-1}g| \geq 1$$

for all real ω.

7.4 FURTHER DISCUSSION OF THE POLE-POSITIONING PROBLEM

In this section, we return to the discussion of the pole-positioning problem for single-input systems. We suppose that we are given the system

$$\dot{x} = Fx + gu \tag{7.4-1}$$

and we desire a control law

$$u = k'x \tag{7.4-2}$$

which is optimal for the usual sort of quadratic loss function, and such that the eigenvalues of $F + gk'$ take on desired values, at least approximately. The approach we adopt here is based on application of the inequality

$$|1 - k'(j\omega I - F)^{-1}g| \geq 1 \tag{7.4-3}$$

for all real ω, which we know is necessary and sufficient for $u = k'x$ to be optimal, given that (1) the pair $[F, g]$ is completely controllable, and (2) the closed-loop system $\dot{x} = (F + gk')x$ is asymptotically stable.

We shall also relate our discussion here to that given earlier in Chapter 5, Sec. 5.4, where we discussed the achieving of desired poles by taking very large ρ in the performance index $\int_{t_0}^{\infty} [u^2 + \rho(d'x)^2]\,dt$, and appropriately choosing d.

First, we find an equivalent statement to (7.4-3). Observe that $\det [sI - (F + gk')] = \det \{[sI - F][I - (sI - F)^{-1}gk']\} = \det [sI - F] \times \det [I - (sI - F)^{-1}gk'] = \det [sI - F][1 - k'(sI - F)^{-1}g]$. [Here, we have used the relations, see Appendix A, $\det (AB) = \det A \det B$ and $\det (I_m + AB) = \det (I_n + BA)$, where the matrices have appropriate dimensions.] Consequently, (7.4-3) and the definitions

$$\psi(s) = \det (sI - F) \quad \text{and} \quad p(s) = \det (sI - F - gk')$$

imply that

$$\frac{p(-j\omega)p(j\omega)}{\psi(-j\omega)\psi(j\omega)} \geq 1$$

or that

$$p(-j\omega)p(j\omega) \geq \psi(-j\omega)\psi(j\omega) \tag{7.4-4}$$

for all real ω. Notice also that Eq. (7.4-4) implies (7.4-3).

Now suppose that a list of n desired closed-loop pole positions is given for (7.4-1). These pole positions determine uniquely the characteristic polynomial $p(s) = s^n + (a_n - k_n)s^{n-1} + \cdots + (a_1 - k_1)$, and, since (7.4-4) implies (7.4-3), these closed-loop pole positions are achievable with an optimal control law $u = k'x$ if and only if (7.4-4) is satisfied. In other words, (7.4-4) provides a test that will decide whether a set of n prescribed closed-loop poles are optimal or not. Moreover, if they are optimal, it is possible to determine an associated control law and performance index using the ideas of the previous section. The pole-positioning problem is thereby solved. Note that in the earlier treatment, the positions of at most $(n - 1)$ poles were prescribed (with the remainder lying well in the left half-plane). By contrast, n pole positions are prescribed here.

What, though, if the list of n poles is not such that (7.4-4) holds? It is obviously not possible to choose an optimal control law generating these desired poles, and some concession must be made. One such concession is to relax the positioning constraint on one real pole (say), and permit its position to be adjustable. Then, of course, we are returning to a situation similar to that discussed earlier, in Sec. 5.4. But the use of (7.4-4) can give us additional insight into the problem. Specifying $(n - 1)$ poles and leaving one unspecified on the real axis means specifying that

$$p(j\omega) = (j\omega + \alpha)p_1(j\omega) \tag{7.4-5}$$

where α is adjustable, whereas $p_1(s)$ has zeros which are the $(n - 1)$ specified poles. Then (7.4-4) may be written

$$(\omega^2 + \alpha^2)p_1(-j\omega)p_1(j\omega) \geq \psi(-j\omega)\psi(j\omega). \tag{7.4-6}$$

The coefficients of ω^{2n} on both sides are the same, and it is therefore possible to find a suitably large value of α such that the inequality holds, irrespective of the polynomials $\psi(s)$ and $p_1(s)$, although the actual value of α depends, of course, on these polynomials. For this and larger values of α (corresponding to a closed-loop pole at or to the left of a certain point on the negative real axis), the closed-loop system defined via (7.4-5) is optimal.

In summary, if a set of n desired closed-loop poles fails to satisfy the inequality criterion, adjustment of one pole (by moving it in the negative direction along the negative real axis sufficiently far) will result in an optimal set of poles, i.e., a set such that the inequality (7.4-4) is satisfied.

We recall that the method of Sec. 5.4, also yielded a set of closed-loop poles among which one or more were a substantial distance from the origin.

With both schemes, a certain difficulty may arise. With one or more poles well in the left half-plane, one or more of the entries k_i of the feedback gain vector must be large. [This follows, for example, by noting that the sum of the closed-loop poles is $-(a_n - k_n)$, the product is $(-1)^n(a_1 - k_1)$, etc. Consequently, the magnitude of the control applied to system (7.4-1) may be large—large enough, in fact, to cause saturation. This conclusion is also borne out by the fact that making ρ large in the performance index $\int_{t_0}^{\infty} [u^2 + \rho(d'x)^2]\, dt$ serves to penalize nonzero states rather than nonzero control. Therefore, any pole-positioning scheme should only be used with caution.

Finally, given $p(s)$ and $\psi(s)$, we note that the actual computations required to check whether the inequality (7.4-4) holds are straightforward. Of course, the problem of checking (7.4-4) is equivalent to one of checking whether a polynomial $r(\cdot)$ in ω^2 is nonnegative for all ω. Procedures based on Sturm's theorem or the equivalent Routh test (see [11] and [12], respectively) provide a check in a finite number of steps involving simple manipulations of the coefficients of $r(\cdot)$. For a survey of other tests, see [13].

Problem 7.4-1. Consider the system

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

Show that it is not possible to choose an optimal control law such that the closed-loop poles are both at $s = -\frac{1}{2}$. Adjust the closed-loop poles until an optimal design is achievable.

Problem 7.4-2. For the same system as Problem 7.4-1, show that it is never possible to obtain complex closed-loop poles which are also optimal for some quadratic performance index.

Problem 7.4-3. Consider the system

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

Show that closed-loop poles which are optimal for a quadratic performance index must have a damping ratio of at least $1/\sqrt{2}$.

Problem 7.4-4. Suppose that the control law $u = k'x$ is optimal for the completely controllable system $\dot{x} = Fx + gu$, and the closed-loop system is asymptotically stable. Show that

$$|\Pi[\lambda_i(F)]| \leq |\Pi[\lambda_i(F + gk')]|.$$

REFERENCES

[1] Perkins, W. R., and J. B. Cruz, Jr., "The Parameter Variation Problem in State Feedback Control Systems," *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 87, March 1965, pp. 120–124.

[2] Kreindler, E., "On the Definition and Application of the Sensitivity Function," *J. Franklin Inst.*, Vol. 285, No. 1, January 1968.

[3] Newcomb, R. W., and B. D. O. Anderson, "A Distributional Approach to Time-Varying Sensitivity," *SIAM J. Applied Math.*, Vol. 15, No. 4, July 1967, pp. 1001–1010.

[4] Newcomb, R. W., *Linear Multiport Synthesis*, McGraw-Hill Book Company, New York, 1966.

[5] Kreindler, E., "Closed-Loop Sensitivity Reduction of Linear Optimal Control Systems," *IEEE Trans. Auto. Control*, Vol. AC-13, No. 3, June 1968, pp. 254–262.

[6] Kalman, R. E., "When is a Linear Control System Optimal?" *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 86, March 1964, pp. 1–10.

[7] Anderson, B. D. O., "The Inverse Problem of Optimal Control," *Rep. No. SEL-66-038* (*TR No. 6560-3*), May 1966, Stanford Electronics Laboratories, Stanford, Calif.

[8] Anderson, B. D. O., "Sensitivity Improvement Using Optimal Design," *Proc. IEE*, Vol. 113, No. 6, June 1966, pp. 1084–1086.

[9] Sobral, M., "Sensitivity in Optimal Control Systems," *Proc. IEEE*, Vol. 56, No. 10, October 1968, pp. 1644–1652.

[10] Sage, A. P., *Optimal Systems Control*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968.

[11] Guillemin, E. A., *Synthesis of Passive Networks*, John Wiley & Sons, Inc., New York, 1957.

[12] Šiljak, D., "Analytic Test for Absolute Stability," *Electron. Letters*, Vol. 4, No. 17, August 1968, pp. 358–359.

[13] Anderson, B. D. O., "Tests for Positivity of a Prescribed Polynomial" (to be published).

CHAPTER 8

STATE ESTIMATOR DESIGN

8.1 THE NATURE OF THE STATE ESTIMATION PROBLEM

The implementation of the optimal control laws discussed hitherto depends on the states of the controlled plant being available for measurement. Frequently, in practical situations, this will not be the case, and some artifice to get around this problem is required. Before outlining the various approaches that may be used, let us dismiss upon practical grounds one theoretically attractive state estimation procedure.

Given a completely observable system, the state vector may be constructed from linear combinations of the output, input, and derivative of these quantities, as we now show. Consider for simplicity a single-output system: thus,

$$\dot{x} = Fx + Gu \tag{8.1-1}$$

$$y = h'x. \tag{8.1-2}$$

Differentiation of (8.1-2) and substitution for $\dot{x}$ from (8.1-1) leads to

$$\dot{y} - h'Gu = h'Fx. \tag{8.1-3}$$

Differentiating again, and substituting again for $\dot{x}$, we get

$$\ddot{y} - h'g\dot{u} - h'Fgu = h'F^2x. \tag{8.1-4}$$

Continuing in this way, we can obtain a set of equations which may be summed

up as

$$z = \begin{bmatrix} h' \\ h'F \\ \cdot \\ \cdot \\ \cdot \\ h'F^{n-1} \end{bmatrix} x$$

or

$$z' = x'[h \quad F'h \quad \cdots \quad (F')^{n-1}h] \tag{8.1-5}$$

with z' a row vector, the entries of which consist of linear combinations of y, u, and their derivatives.

Now one of the results concerning complete observability is that the pair $[F, H]$ is completely observable if and only if the matrix $[H \quad F'H \cdots (F')^{n-1}H]$ has rank n. (A demonstration that this result follows from one of the earlier definitions given for complete observability is requested in Problem 8.1-1. See also Appendix B.) In the scalar output case, the matrix $[h \quad F'h \cdots (F')^{n-1}h]$ becomes square, and thus has rank n if and only if it is nonsingular.

Therefore, with the system of (8.1-1) and (8.1-2) completely observable, Eq. (8.1-5) implies that the entries of the state vector x are expressible as linear combinations of the system output, input, and their derivatives.

From the strictly theoretical point of view, this idea for state estimation works. From the practical point of view, it will not. The reason, of course, is that the presence of noise in u and y will lead to vast errors in the computation of x, because differentiation of u and y (and, therefore, of the associated noise) is involved. This approach must thus be abandoned on practical grounds.

However, it may be possible to use a slight modification of the preceding idea, because it is possible to build "approximate" differentiators that may magnify noise but not in an unbounded fashion. The rate gyro, normally viewed as having a transfer function of the form Ks, in reality has a transfer function of the form $Kas/(s + a)$, where a is a large constant. Problem 8.1-2 asks the student to discuss how this transfer function might approximately differentiate.

Generally speaking, a less ad hoc approach must be adopted to state estimation. Let us start by stating two desirable properties of a state estimator:

1. It should be a system of the form of Fig. 8.1-1, with inputs consisting of the system input and output, and output $x_e(t)$ at time t, an *on-line* estimate of $x(t)$. This should allow implementation of the (optimal) control law $u = K'x_e$, according to the scheme of Fig. 8.1-2.

Fig. 8.1-1 Desired structure of estimator.

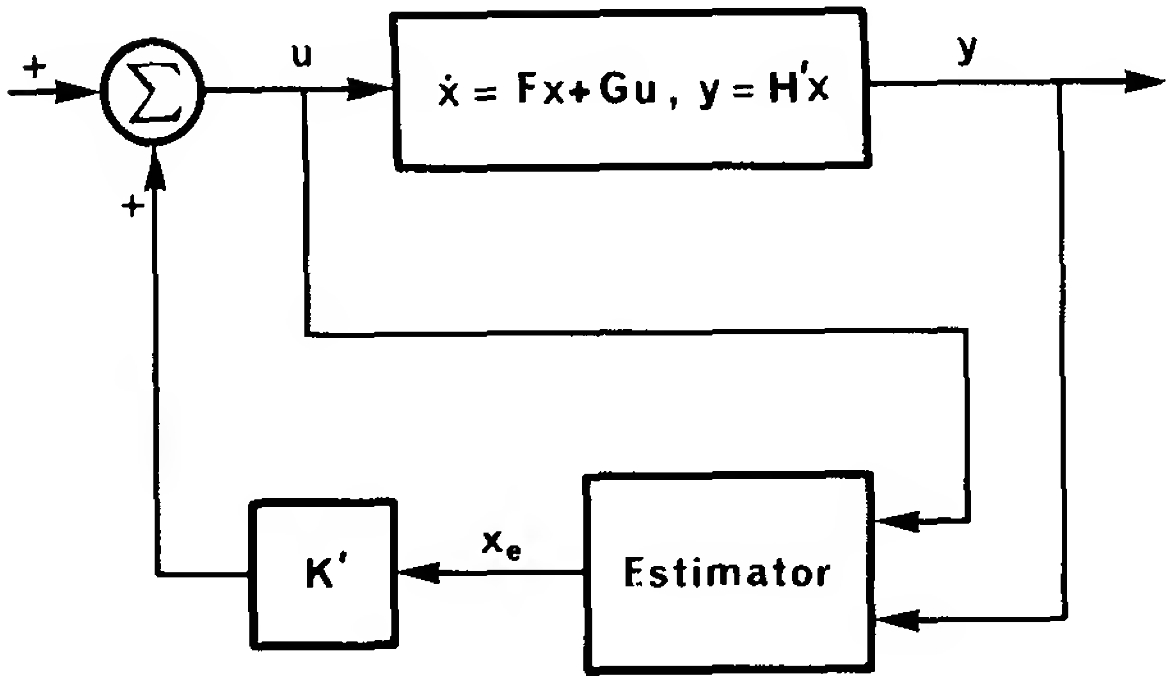


Fig. 8.1-2 Use of estimator in implementing a control law.

2. It should function in the presence of noise. Preferably, it should be possible to optimize the action of the estimator in a noisy environment—i.e., to ensure that the noise has the least possible effect when the estimator is used in connection with controlling a system.

As we shall show, these properties are both more or less achievable. Estimators designed to behave *optimally in the presence of noise* turn out to consist of linear, finite-dimensional systems, if the system whose states are being estimated is linear and finite-dimensional. Moreover, if the dimensions of the system whose states are being estimated and the estimator are the same, and if the system whose states are being estimated is time invariant and the associated noise is stationary, the estimator is time invariant. Also, all these properties hold irrespective of the number of inputs and outputs of the system whose states are being estimated.

A further interesting property of such estimators is that their design is independent of the associated optimal feedback law K' shown in Fig. 8.1-2, or of any performance index used to determine K'. Likewise, the determination of K' is independent of the presence of noise, and certainly independent of the particular noise statistics. Yet, use of the control law $u = K'x_e$ turns out to be not merely approximately optimal, but exactly optimal for a modified form of performance index. This point is discussed at length in Chapter 9, Sec. 9.4.

If satisfactory rather than optimal performance in the presence of noise is acceptable, two simplifications can be made. First, the actual computational procedure for designing an estimator (at least for scalar output systems) becomes far less complex. Second, it is possible to simplify the structure of the estimator, if desired. In other words, the designer can stick with the same estimator structure as is used for optimal estimators (although optimality

is now lost), or he may opt for a simpler structure. In general, the estimator with simpler structure is more noise prone. The nature of the structural simplification is that the dimension of the estimator is lowered, by one for a single-output system and sometimes by a larger number for a multiple-output system.

Because they are simpler to understand, we shall discuss first those estimators that are not designed on the basis of optimizing their behavior in the presence of noise; this class of estimators may be divided into two subclasses, consisting of those estimators with the same dimension as the system and those with lower dimension. Then the class of estimators offering optimal noise performance will be discussed. The next chapter will consider the use of estimators in implementing control laws, and the various properties of the associated plant controller arrangement.

One point to keep in mind is that the output of the estimator at time t, $x_e(t)$, is normally an estimate rather than an exact replica of the system state $x(t)$, even when there is no noise present. In other words, no estimator is normally perfectly accurate, and thus feedback laws using an estimate rather than the true states will only approximate the ideal situation. Despite this, little or no practical difficulty arises in controller design as a result of the approximation.

Problem 8.1-1. We have earlier adopted the following definition: $[F, H]$ is completely observable if and only if $H'e^{Ft}x_0 = 0$ for all t implies $x_0 = 0$. Show that a necessary and sufficient condition for complete observability according to this definition is that rank $[H \quad F'H \cdots (F')^{n-1}H] = 0$. {*Hint:* Suppose that $H'e^{Ft}x_0 = 0$ for all t and some nonzero x_0. Show by differentiating and choosing a specific t that $x_0'[H \quad F'H \cdots (F')^{n-1}H] = 0$. For the converse, suppose that $x_0'[H \quad F'H \cdots (F')^{n-1}H] = 0$ for some nonzero x_0, and use the fact that $e^{Ft} = \alpha_0(t)I + \alpha_1(t)F + \cdots + \alpha_{n-1}(t)F^{n-1}$ for certain scalar functions $\alpha_i(t)$, to conclude that $H'e^{Ft}x_0 = 0$ for all t.}

Problem 8.1-2. In what sense can the transfer function Ks be approximated by the transfer function $Kas/(s + a)$? What is the significance of the size of a? Discuss the effect of the noise with the two transfer functions (e.g., evaluate the output spectra for one or two types of input noise spectra).

Problem 8.1-3. Show that the circuit of Fig. 8.1-3 can be used as an approximate differentiator. Discuss the effect of noise in e_i.

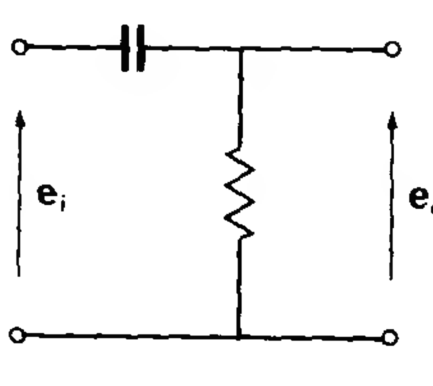


Fig. 8.1-3 Circuit for Problem 8.1-3.

8.2 NONSTATISTICAL ESTIMATOR DESIGN WITH FULL ESTIMATOR DIMENSION

In this section, we shall consider the design of estimators for systems of the form

$$\dot{x} = Fx + Gu \tag{8.2-1}$$

$$y = H'x. \tag{8.2-2}$$

We shall assume that F, G, and H are time invariant, although it is possible to extend the theory to time-varying systems. The estimators will be of the general form

$$\dot{x}_e = F_e x_e + G_{1e}u + G_{2e}y \tag{8.2-3}$$

with F_e of the same size as F. Equation (8.2-3) reflects the fact that the inputs to the estimator are the input and output associated with Eqs. (8.2-1) and (8.2-2), hereafter called the plant input and output. The estimator itself is a linear, finite-dimensional, time-invariant system, the output of which is the estimated state vector of the system.

Before we give a detailed statement specifying how to choose F_e, G_{1e}, and G_{2e}, it is worthwhile to make two helpful observations:

1. It would be futile to think of constructing an estimator using (8.2-3) if the plant equations (8.2-1) and (8.2-2) did not define a completely observable pair $[F, H]$, because lack of complete observability implies the impossibility of determining, by any technique at all, the state of the plant from the plant input and output.
2. One might expect the estimator to be a model for the plant, for suppose that at some time t_0, the plant state $x(t_0)$ and estimator state $x_e(t_0)$ were the same. Then the way to ensure that at some subsequent time $x_e(t)$ will be the same as $x(t)$ is to require that the estimator, in fact, model the plant—i.e., that

$$\dot{x}_e = Fx_e + Gu. \tag{8.2-4}$$

Clearly, though, Eq. (8.2-4) is not satisfactory if $x_e(t_0)$ and $x(t_0)$ are different. What is required is some modification of (8.2-4) that reduces to (8.2-4) if $x_e(t_0)$ and $x(t_0)$ are the same, and otherwise tends to shrink the difference between $x_e(t)$ and $x(t)$ until they are effectively the same. Now we may ask what measure we could physically construct of the difference between $x_e(t)$ and $x(t)$. Certainly there is no direct measure, but we do have available $H'x$, and therefore we could physically construct $H'[x_e(t) - x(t)]$. Hopefully, the complete observability of the plant would ensure that, over a nonzero interval of time, this quantity contained as much information as $x_e(t) - x(t)$.

These considerations suggest that instead of (8.2-4), we might aim for

$$\dot{x}_e = Fx_e + Gu + K_eH'[x_e - x] \tag{8.2-5}$$

as an estimator equation. This scheme is shown in Fig. 8.2-1. The equation has the property that if x and x_e are the same at some time t_0, then they will be the same for all $t \geq t_0$, the third term on the right-hand side being zero for all t. Furthermore, it has the property that possible judicious selection of K_e—i.e., judicious introduction of a signal into the estimator reflecting the difference between $H'x_e(t)$ and $y(t) = H'x(t)$—may ensure that the error between $x_e(t)$ and $x(t)$ becomes smaller as time advances. Let us now check this latter possibility.

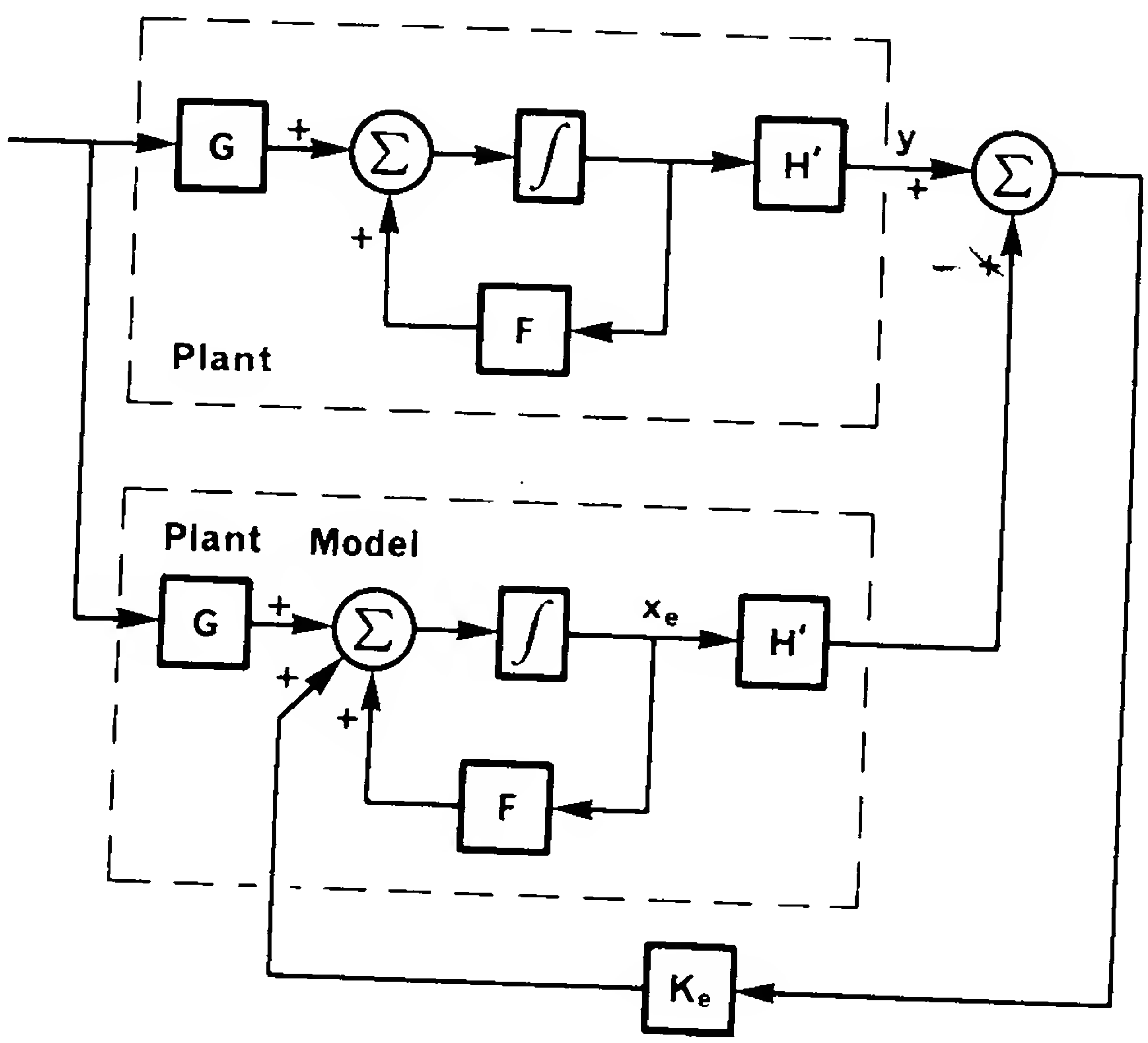


Fig. 8.2-1 Estimator, illustrating plant model concept.

Subtracting (8.2-5) from (8.2-1), we find that

$$\frac{d}{dt}(x - x_e) = F(x - x_e) + K_eH'(x - x_e)$$

$$= (F + K_eH')(x - x_e) \tag{8.2-6}$$

follows. Consequently, if the eigenvalues of $(F + K_eH')$ have negative real parts, $x - x_e$ approaches zero at a certain exponential rate, and $x_e(t)$ will effectively track $x(t)$ after a time interval determined by the eigenvalues of $F + K_eH'$.

Let us recapitulate. We postulated the availability of the input and output of the plant defined by (8.2-1) and (8.2-2), together with complete observability of the plant. By rough arguments, we were led to examining the possi-

bility of an estimator design of the form of Eq. (8.2-5), which may be rewritten as

$$\dot{x}_e = (F + K_eH')x_e + Gu - K_ey. \tag{8.2-7}$$

[Figure 8.2-2 shows how this equation can be implemented.] Then we were able to conclude that if K_e could be chosen so that the eigenvalues of $F + K_eH'$ had negative real parts, Equation (8.2-7) did, in fact, specify an estimator, in the sense that $x_e - x$ approaches zero at some exponential rate. Note that at this stage, we have not considered any questions relating to the introduction of noise.

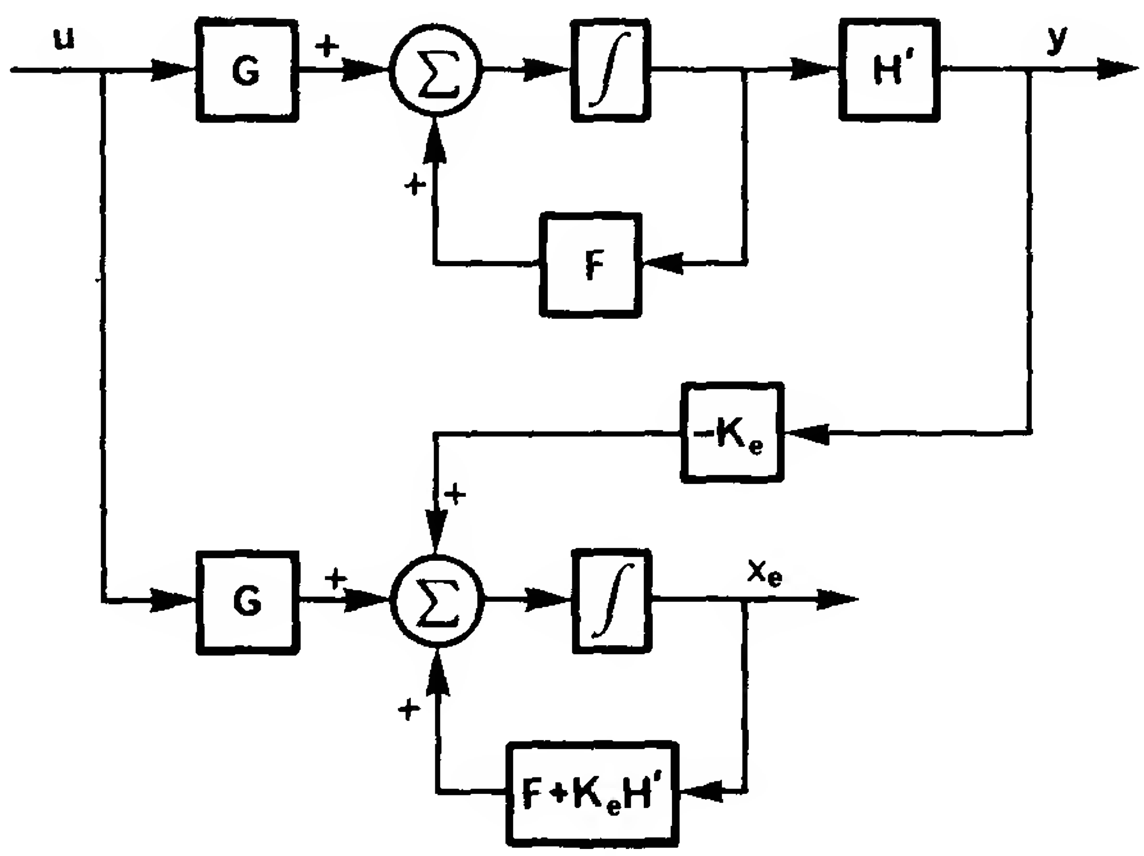


Fig. 8.2-2 Estimator with minor simplification.

The natural question now arises: When can K_e be found so that $F + K_eH'$ has arbitrary eigenvalues? The answer is—precisely when the pair $[F, H]$ is completely observable! (The earliest statement of this result appears to be in [1], but it or an equivalent result has since received wide exposure—e.g., [2] through [4].) Therefore, aside from the computational details involved in determining K_e, and aside from checking the noise performance, we have indicated one solution to the estimator problem.

The scheme of Fig. 8.2-1, earlier regarded as tentative, has now been shown to constitute a valid estimator. The estimator is a model of the plant, with the addition of a driving term reflecting the error between the plant output $y = H'x$ and the variable $H'x_e$, which has the effect of causing x_e to approach x. Figure 8.2-2 shows an alternative valid estimator representation equivalent to that shown in Fig. 8.2-1. In this second figure, however, the concept of the estimator as a model of the plant becomes somewhat submerged.

Let us now consider the question of the effect of noise on estimator operation. If noise is associated with u and y, then inevitably it will be smoothed by the estimator; and if white noise is associated with either u or

y (i.e., noise with a uniform power spectrum), the spectrum of the noise in x_e will fall away at high frequencies. In general, the amount of output noise will depend on the choice of K_e, but the problem associated with passing noise into a differentiator will never be encountered. The choice of K_e also affects the rate at which x_e approaches x, because this rate is, in turn, governed by the eigenvalues of $F + K_eH'$. It might be thought that the best way to choose K_e would be to ensure that the eigenvalues of $F + K_eH'$ had real parts as negative as possible, so that the approach of x_e to x would be as rapid as possible. This is so, with one proviso. As the eigenvalues of $F + K_eH'$ get further into the left half-plane, the effective bandwidth of the estimator could be expected to increase, and, accordingly, the noise in x_e due to noise in u and y could be expected to increase. Therefore, noise will set an upper limit on the speed with which x_e might approach x. The situation with using differentiation to estimate x is that the noise becomes infinite, with the estimation time infinitesimally small. The use of estimators of the form just described is aimed at trading off speed of estimation against minimization of loss of performance due to noise. The optimal estimators of Sec. 8.4 essentially achieve the best compromise.

In the remainder of this section, we shall discuss the problem of computing K_e. In the next chapter, we shall return to the investigation of the performance of the estimator in an optimal control system, and to minor variants of it.

We shall first discuss the question of computing K_e for a single-output system. In this case, K_e becomes a vector, and will accordingly be denoted by k_e. The problem of finding k_e to produce arbitrary eigenvalues for the matrix $F + k_eh'$ is easy if F and h have the following form:

$$F = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -a_1 \\ 1 & 0 & \cdot & \cdot & 0 & -a_2 \\ 0 & 1 & \cdot & \cdot & 0 & -a_3 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -a_n \end{bmatrix} \qquad h = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}, \tag{8.2-8}$$

for if $k_e' = [k_{1e} \quad k_{2e} \quad \cdots \quad k_{ne}]$, it follows that

$$F + k_eh' = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -(a_1 - k_{1e}) \\ 1 & 0 & \cdot & \cdot & 0 & -(a_2 - k_{2e}) \\ 0 & 1 & \cdot & \cdot & 0 & -(a_3 - k_{3e}) \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -(a_n - k_{ne}) \end{bmatrix} \tag{8.2-9}$$

and

$$\det[sI - (F + k_e h')] = s^n + (a_n - k_{ne})s^{n-1} + \cdots + (a_1 - k_{1e}), \tag{8.2-10}$$

as may easily be shown. Now specification of the eigenvalues of $F + k_e h'$ is equivalent to specification of the characteristic polynomial of $F + k_e h'$, or, equivalently, the coefficients $a_i - k_{ie}$ in (8.2-10). Therefore, with the a_i known and the coefficients $a_i - k_{ie}$ specified, the vector k_e is immediately determined.

When F and h are not in the form just shown, it is necessary to transform them into this form via an appropriate basis change, to choose a k_e in this new basis, and then to determine the k_e in the old basis. In precise terms, we define the matrix

$$T = \begin{bmatrix} a_2 & a_3 & \cdot & \cdot & a_n & 1 \\ a_3 & a_4 & \cdot & a_n & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & a_n & \cdot & \cdot & \cdot & \cdot \\ a_n & 1 & 0 & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix} \begin{bmatrix} h' \\ h'F \\ h'F^2 \\ \cdot \\ \cdot \\ h'F^{n-1} \end{bmatrix} \tag{8.2-11}$$

where $s^n + a_n s^{n-1} + \cdots + a_1$ is the characteristic polynomial of F.

Recall that the matrix $[h \quad F'h \quad \cdots \quad (F')^{n-1}h]$ has rank n if and only if $[F, h]$ is completely observable. Therefore, the matrix T of (8.2-11) is nonsingular if and only if $[F, h]$ is completely observable.

The significance of T is that, as we shall show,

$$\hat{F} = TFT^{-1} = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -a_1 \\ 1 & 0 & \cdot & \cdot & 0 & -a_2 \\ 0 & 1 & \cdot & \cdot & 0 & -a_3 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -a_n \end{bmatrix} \qquad \hat{h} = (T^{-1})'h = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}. \tag{8.2-12}$$

Accordingly, if $\hat{k}_e$ is determined so that $\hat{F} + \hat{k}_e \hat{h}'$ has desired eigenvalues (as is easy to do), it follows that with

$$k_e = T^{-1}\hat{k}_e \tag{8.2-13}$$

the matrix $F + k_e h'$ has the same eigenvalues.

To verify the claim concerning the structure of $\hat{F}$ and $\hat{h}$, we follow an idea of Tuel [5] applied originally to completely controllable systems. Define

the row vectors $t'_1, t'_2, \ldots, t'_n$ sequentially through

$$t'_n = h'$$
$$t'_i = t'_{i+1}F + a_{i+1}h' \qquad i < n. \tag{8.2-14}$$

Observe that

$$\begin{aligned} t'_1F &= t'_2F^2 + a_2h'F \\ &= t'_3F^3 + a_3h'F^2 + a_2h'F \\ &\cdot \\ &\cdot \\ &\cdot \\ &= t'_nF^n + a_{n-1}h'F^{n-1} + \cdots + a_2h'F \\ &= h'[F^n + a_{n-1}F^{n-1} + \cdots + a_2F] \end{aligned}$$

or

$$t'_1F = -a_1h' \tag{8.2-15}$$

on using the fact that F satisfies its characteristic equation. Now,

$$\begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -a_1 \\ 1 & 0 & \cdot & \cdot & 0 & -a_2 \\ 0 & 1 & \cdot & \cdot & 0 & -a_3 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -a_n \end{bmatrix} \begin{bmatrix} t'_1 \\ t'_2 \\ t'_3 \\ \cdot \\ \cdot \\ t'_n \end{bmatrix} = \begin{bmatrix} -a_1t'_n \\ t'_1 - a_2t'_n \\ t'_2 - a_3t'_n \\ \cdot \\ \cdot \\ t'_{n-1} - a_nt'_n \end{bmatrix} = \begin{bmatrix} t'_1F \\ t'_2F \\ t'_3F \\ \cdot \\ \cdot \\ t'_nF \end{bmatrix} = \begin{bmatrix} t'_1 \\ t'_2 \\ t'_3 \\ \cdot \\ \cdot \\ t'_n \end{bmatrix} F$$

where we have made use of Eqs. (8.2-14) and (8.2-15). Also,

$$[0 \quad 0 \quad \cdot \quad \cdot \quad 1] \begin{bmatrix} t'_1 \\ t'_2 \\ \cdot \\ \cdot \\ t'_n \end{bmatrix} = t'_n = h'.$$

Therefore, with

$$T = \begin{bmatrix} t'_1 \\ t'_2 \\ \cdot \\ \cdot \\ t'_n \end{bmatrix}, \tag{8.2-16}$$

Eqs. (8.2-12) follow, with $\hat{F}$ and $\hat{h}$ possessing the desired structure. Using the definitions (8.2-14), it is also straightforward to verify that T has the form of Eq. (8.2-11).

Let us now consider a simple example. Suppose we are given the plant equations

$$\dot{x} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u \qquad (8.2\text{-}17)$$
$$y = [2 \quad -1]\, x$$

which describe a second-order system with transfer function $(s + 1)/s(s - 1)$. We shall provide an estimator design such that the eigenvalues of $F + k_e h'$ are both at -10. Then $x - x_e$ will decay as fast as e^{-10t}. First, the matrix T must be found which will take F and h into the special format of Eq. (8.2-12). We have $\det(sI - F) = s^2 - s$, and therefore, from Eq. (8.2-11),

$$T = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}.$$

Also,

$$T^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix}.$$

Then

$$\hat{F} = TFT^{-1} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \qquad \hat{h} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and therefore,

$$\hat{k}_e = \begin{bmatrix} -100 \\ -\ 21 \end{bmatrix}$$

so that

$$\hat{F} + \hat{k}_e \hat{h}' = \begin{bmatrix} 0 & -100 \\ 1 & -20 \end{bmatrix}$$

has two eigenvalues of -10. Then,

$$k_e = T^{-1}\hat{k}_e = \begin{bmatrix} -\ 60.5 \\ -100 \end{bmatrix}.$$

Using (8.2-7), we find the estimator equation to be

$$\dot{x}_e = \begin{bmatrix} -120 & 60.5 \\ -200 & 100 \end{bmatrix} x_e + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u + \begin{bmatrix} 60.5 \\ 100 \end{bmatrix} y.$$

One interesting conclusion from this equation is that apparently noise associated with the output measurements y will have a significantly greater effect in terms of introducing noise into x_e than will noise in the input measurement u.

We now turn to a discussion of estimator design for multiple-output plants. The computational procedure leading to selection of a matrix K_e such that $F + K_eH'$ has arbitrary eigenvalues for a multiple output plant is exceedingly involved, and will not be studied in generality here. Discussion may be found in [1] through [4] and [6] through [8].

However, there are two classes of multiple output systems where the preceding theory is sufficient. The first class consists of those multiple-output systems for which the state vector is observable from one output alone—i.e., if h_i is the ith column of the H matrix, $[F, h_i]$ is completely observable for one i. Observer design then can proceed using the scalar output variable $h'x$ and disregarding the other components of the output. One would expect this to be far from optimal in a noisy situation, since information is being thrown away in estimating the state variable.

The second class of plants are those made up of a cascade of small plants. Frequently in practice, various elements of an overall system may be individually distinguishable, and the system will have the general form of Fig. 8.2-3, where for convenience we have indicated the individual small plants by transfer functions $w_i(s)$ rather than by state-space equations. The overall plant transfer function is $w(s) = \prod_i w_i(s)$.

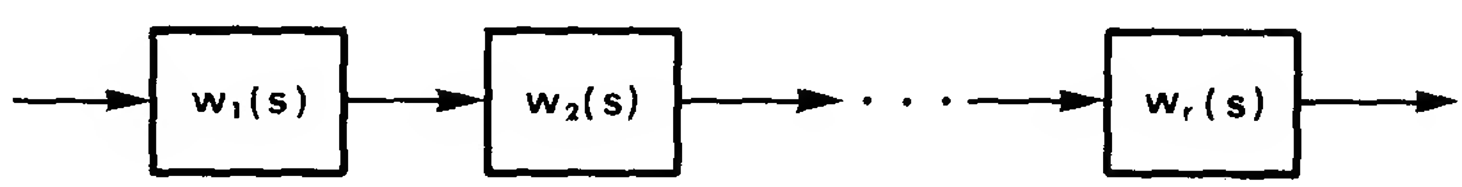


Fig. 8.2-3 Cascade of small plants providing a large plant.

One way to carry out state estimation in this case is to use a number of small estimators, call them $E_1, E_2, \ldots, E_r$, with estimator E_i estimating the states of the small plant with transfer function $w_i(s)$. Estimator E_i is, of course, driven by the input and output of $w_i(s)$, and will have the same dimension (when viewed as a linear system) as does w_i. Figure 8.2-4 shows the composite estimation scheme. Evidently, the dimension of the overall estimator is the sum of the dimensions of the individual $w_i(s)$ plants.

It may be felt that there is nothing to gain by using the scheme of Fig. 8.2-4 instead of that in Fig. 8.2-5, where the overall plant input and output alone are used to drive an estimator. After all, nothing is saved in terms of the overall estimator dimension by using the scheme of Fig. 8.2-4. There are actually several reasons. First, one or more of the $w_i(s)$ may be such that there is a pole-zero cancellation between two or more of the $w_i(s)$. In this instance, it may not be true that measurement of the overall plant output alone will allow observation of all the states—i.e., if a composite F matrix and h vector are formed for the whole plant, the pair $[F, h]$ may not be completely observable. Second, the noise properties associated with the first scheme may be better, if only because more information is being used than in the scheme

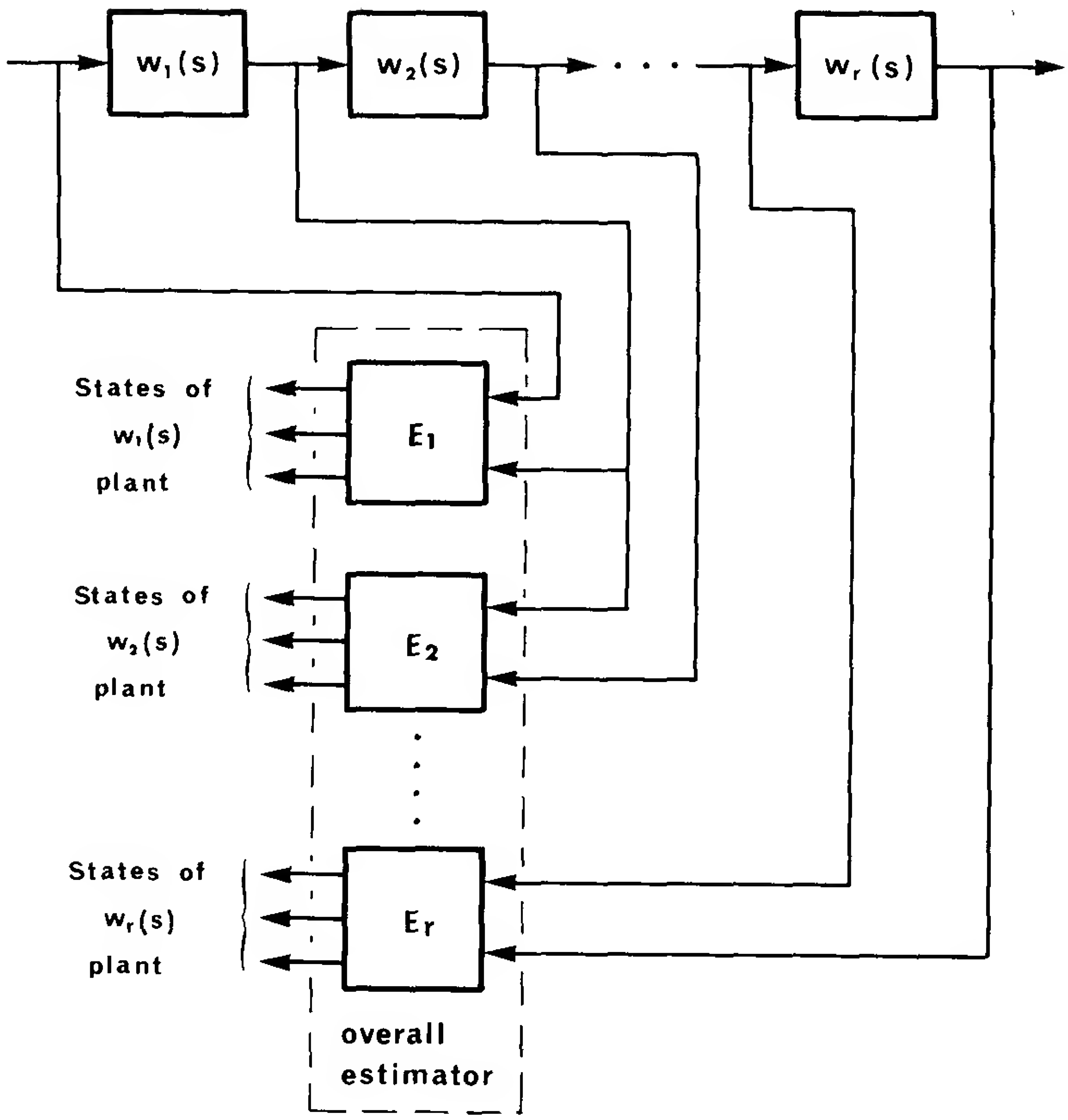


Fig. 8.2-4 Estimator for the scheme of Fig. 8.2-3.

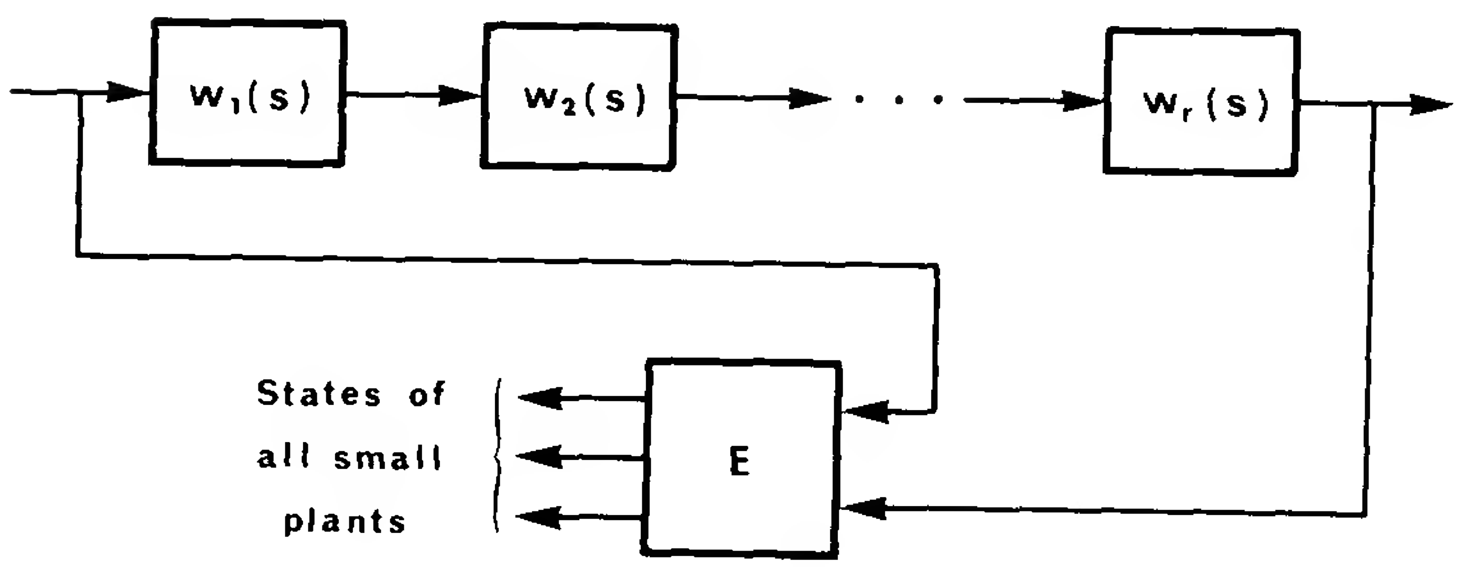


Fig. 8.2-5 Alternative estimation scheme.

of Fig. 8.2-5, where some is actually being discarded. It would be intuitively reasonable to suppose that if there were almost, but not quite, a pole-zero cancellation between two of the small plants, the estimator of Fig. 8.2-5 would be especially susceptible to noise. Third, depending on the physical devices used to implement the estimators, it may be easier to construct the individual low-order estimators of Fig. 8.2-4 than the single high-order one

of Fig. 8.2-5. Fourth, if any of the small plants have transfer functions of the form $\alpha_1/(s + \beta_1)$ or $(s + \alpha_1)/(s + \beta_1)$, estimation of the state can be immediately achieved, since in the case of the first form, the state is the output (or an appropriate multiple), and in the case of the second form, the state is a linear combination of the input and output. Accordingly, there is reduction in the complexity of the associated estimator. Finally, we shall show in the next chapter that if all we really require is $k'x_e$ rather than x_e, in order, say, to implement a feedback law, the scheme of Fig. 8.2-4 allows further drastic simplification in that the set of estimators E_i may be replaced by a single estimator with the same number of inputs as shown, with output $k'x_e$, and with dimension equal to the greatest of the dimensions of the E_i. This is an extremely powerful argument in favor of the estimator arrangement of Fig. 8.2-4 (or, at least, a modification of it) as opposed to the estimator of Fig. 8.2-5.

It is interesting to note that the scheme of [9] for estimation of the state of a multiple-output system relies on representing the system in a certain sense as a cascade of smaller subsystems, the inputs and outputs of which are all available for measurement.

We remarked earlier that it is impossible to estimate all the states of a plant that are not observable. However, in general, it is still possible to estimate the observable components of the states of such a plant; if the unobservable components decay extremely fast, this partial estimation may be adequate. Problem 8.2-3 asks for an examination of estimator design in these circumstances.

Problem 8.2-1: For the system (8.2-17), design a state estimator such that $F + k_e h'$ has two eigenvalues of -5.

Problem 8.2-2: Design an estimator for

$$\dot{x} = \begin{bmatrix} -2 & 1 \\ 0 & -2 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$
$$y = [1 \quad 0]x.$$

Problem 8.2-3: If a system is not completely observable, by a coordinate basis change it may be put into the form

$$\dot{x} = \begin{bmatrix} F_{11} & 0 \\ F_{21} & F_{22} \end{bmatrix} x + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} u$$
$$y = [H_1' \quad 0]x,$$

where $[F_{11}, H_1]$ is completely observable, see Appendix B. Show how to estimate certain of the components of x, and show why it is impossible to estimate the remainder.

Problem 8.2-4: Devise a computational procedure for estimator design for the following system

$$\dot{x} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -3 & 0 & 0 & 0 & 0 \\ 0 & 1 & -3 & 0 & 0 & 0 & 0 \\ 4 & 1 & 9 & 0 & 0 & 0 & -1 \\ 2 & 1 & 3 & 1 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 1 & 0 & -6 \\ 1 & 1 & 1 & 0 & 0 & 1 & -4 \end{bmatrix} x + \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 1 \\ 2 \\ 1 \end{bmatrix} u$$

$$y = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} x.$$

[*Hint:* Examine K_e of the form

$$K_e' = \begin{bmatrix} \alpha & \beta & \gamma & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta & \epsilon & \phi & \zeta \end{bmatrix}$$

where the Greek letters denote elements that are, in general, nonzero.]

Problem 8.2-5: Consider the scheme of Fig. 8.2-6. Write down state-space equations for the individual blocks, and design a state estimator assuming availability of the signals labeled y_1 and y_2. Discuss the situation where u and y_2 alone are available.

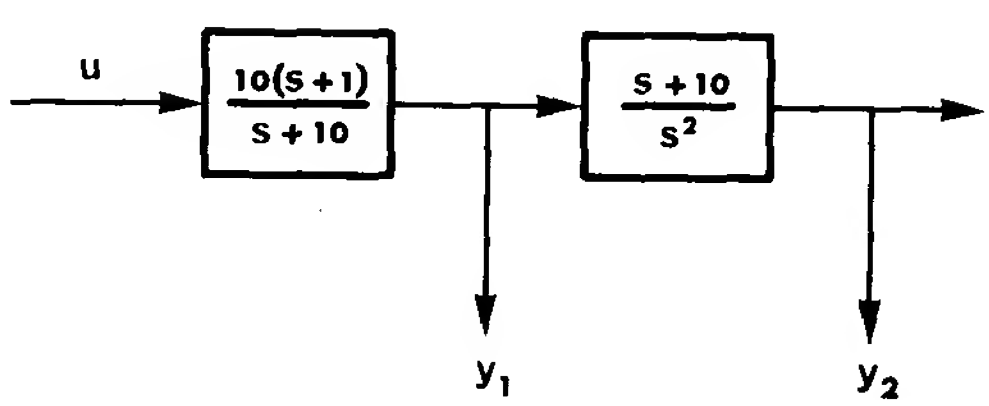


Fig. 8.2-6 System for Problem 8.2-5.

8.3 NONSTATISTICAL ESTIMATOR DESIGN WITH REDUCED ESTIMATOR DYNAMICS

Our aim in this section is similar to that of the previous section, with a few modifications. We assume that a plant is given with the following equations:

$$\dot{x} = Fx + Gu \tag{8.3-1}$$

$$y = H'x \tag{8.3-2}$$

with F, G, and H constant. We seek a system of the form

$$\dot{w} = F_e w + G_{1e}u + G_{2e}y \tag{8.3-3}$$

where F_e has lower dimension than F; the vector $w(t)$ is required to have the property that from it and $y(t)$, a state estimate $x_e(t)$ can be constructed with $x(t) - x_e(t)$ approaching zero at an arbitrarily fast exponential rate. Notice that we have required $x_e(t)$ to depend on the values taken by w and y at time t alone—i.e., we require $x_e(t)$ to be a *memoryless* transformation (indeed, as it turns out, a linear transformation) of $w(t)$ and $y(t)$. The general arrangement is shown in Fig. 8.3-1.

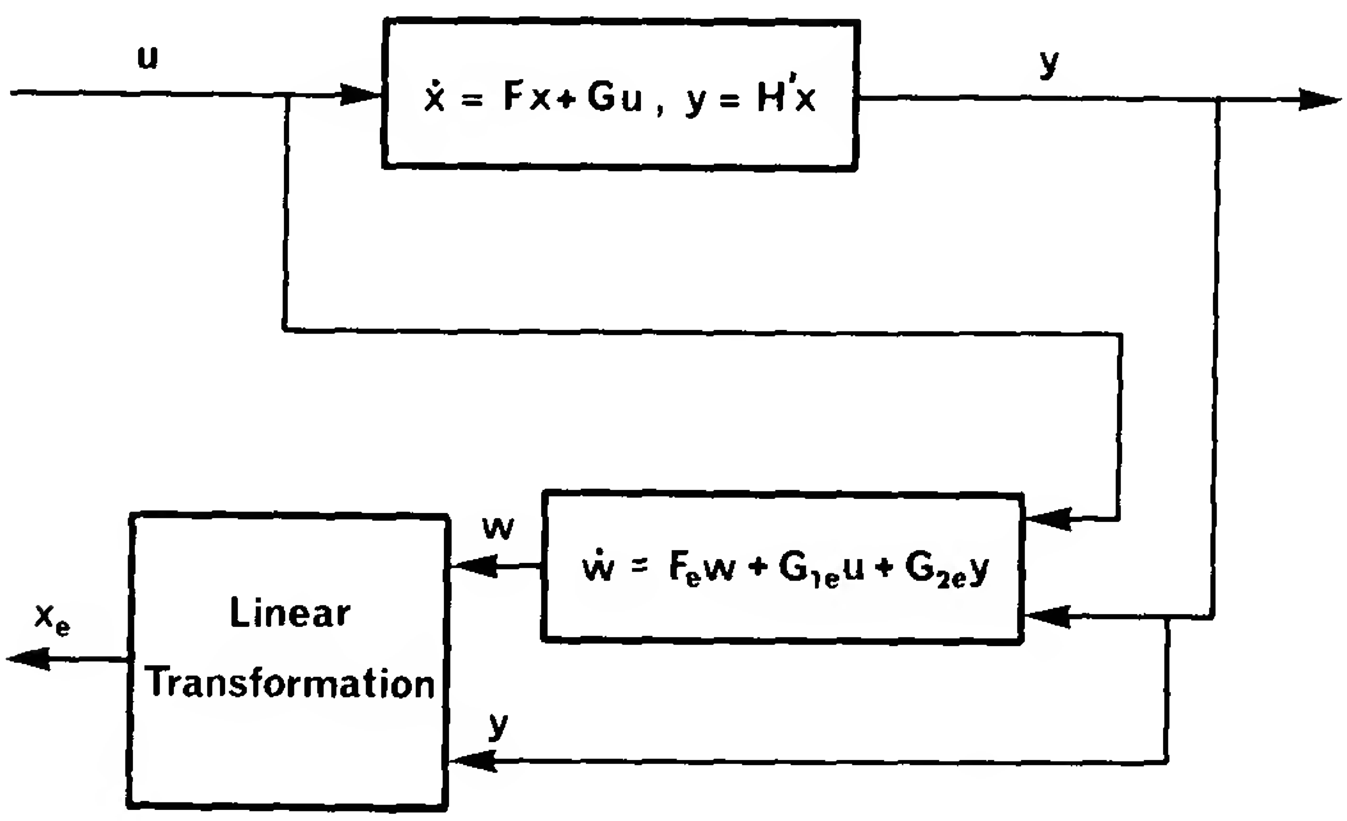


Fig. 8.3-1 General structure of the Luenberger estimator.

Estimators of this variety were first described in [10] by Luenberger for the scalar output case; then the multiple-output case was given in [9]. For single-output systems, the idea suggesting that a reduction in estimator order might be possible is a very simple one: In an appropriate coordinate basis, the plant output constitutes one entry of the state vector. Therefore, if the state vector is an n vector, only $(n - 1)$ components really need be estimated. The generalization for multiple-output systems is also not difficult to understand: If there are r independent outputs, then, plausibly, these can be used to deduce r components of the state vector when an appropriate coordinate basis is used, leaving $(n - r)$ components to be estimated. It then becomes possible, although by no means obvious, to estimate these remaining $(n - r)$ components with an $(n - r)$-dimensional system. The difficulty lies in translating these somewhat imprecise notions into an effective computational scheme.

Most of our discussion will be restricted to single-output systems. Here, if the plant dimension is n, the estimator dimension is $(n - 1)$. As explained in the last section, the assumption that the plant is observable is

vital. Thus, we suppose that (8.3-2) is replaced by $y = h'x$, with $[F, h]$ a completely observable pair.

We shall now deduce some desirable properties of the estimator, following which a computational procedure for estimator design will be given. Without specifically writing down a coordinate basis change, we can conceive a basis change that will have the property that y is the last entry of the state vector x in the new coordinate basis. That is,

$$y = [0 \quad 0 \quad \cdots \quad 0 \quad 1]x. \tag{8.3-4}$$

With an overbar denoting the deletion of the last row of a vector or matrix, we can then write the plant equations as

$$\begin{bmatrix} \dot{\bar{x}} \\ \dot{x}_n \end{bmatrix} = \left[\begin{array}{ccc|c} & F_e & & \bar{b} \\ \hline f_{n1} & \cdots & f_{nn-1} & f_{nn} \end{array}\right] \begin{bmatrix} \bar{x} \\ x_n \end{bmatrix} + \begin{bmatrix} & \bar{G} & \\ g_{n1} & \cdots & g_{nr} \end{bmatrix} u \tag{8.3-5}$$

where the F matrix of the plant has been partitioned conformably with the partitioning of x, and where x_n is actually y. The matrix F_e appearing in (8.3-5) is a submatrix of F, and later appears in the estimator equation.

An estimator of the sort given in the *previous* section would have an equation of the form

$$\begin{bmatrix} \dot{\bar{x}}_e \\ \dot{x}_{ne} \end{bmatrix} = \left[\begin{array}{ccc|c} & F_e & & \bar{b} \\ \hline f_{n1} & \cdots & f_{nn-1} & f_{nn} \end{array}\right] \begin{bmatrix} \bar{x}_e \\ x_{ne} \end{bmatrix} + \begin{bmatrix} & \bar{G} & \\ g_{n1} & \cdots & g_{nr} \end{bmatrix} u + k_e[h'x_e - y] \tag{8.3-6}$$

where k_e is chosen so that $F + k_e h'$ has desired eigenvalues. If, however, x_n is known to be y, there is no point in including a further estimation of x_n using x_{ne}, and the last row of (8.3-6) might just as well be deleted. In making this deletion, observe also what happens to the last term of the right side of (8.3-6): $h'x_e$ is just x_{ne}, which is taken to be y from the start. Thus, this term disappears, and we are left with

$$\dot{\bar{x}}_e = F_e\bar{x}_e + \bar{b}x_{ne} + \bar{G}u = F_e\bar{x}_e + \bar{b}y + \bar{G}u. \tag{8.3-7}$$

In reducing the dimension of the estimator equations, we have unfortunately lost the effect of the feedback term $k_e[h'x_e - y]$ of the full-order estimator (8.3-6), which, it will be recalled, served to define the rate at which the state estimate converged. Equation (8.3-7), the $(n - 1)$th order estimator, has no gain vectors, such as k_e, that might be adjusted to set the rate of convergence. However, the matrix F_e is adjustable, as we shall note subsequently, and we can easily see that this adjustment provides a tool for successful estimator design. Let us consider the convergence properties of the estimator of (8.3-7); deletion of the last row of (8.3-5) yields

$$\dot{\bar{x}} = F_e\bar{x} + \bar{b}y + \bar{G}u. \tag{8.3-8}$$

Subtracting (8.3-7) from (8.3-8) yields

$$\frac{d}{dt}(\bar{x} - x_e) = F_e(\bar{x} - \bar{x}_e). \tag{8.3-9}$$

Consequently, all convergence properties are summed up by the eigenvalues of F_e. Hence, if a coordinate basis is found with the properties that (1) y is identical with the last entry of x, and (2) that the $(n-1) \times (n-1)$ top left submatrix F_e of the F matrix has arbitrary eigenvalues (even though, of course, the eigenvalues of the F matrix are invariant with coordinate basis change), then we can successfully design an estimator.

In the remainder of this section, we shall discuss a computational procedure leading to a coordinate basis with the desired properties. The computational procedure to be presented amounts to revealing an appropriate change of coordinate basis. To illustrate how the change of basis is derived, we shall, in fact, describe two successive basis changes, each of which individually has some significance. In actual practice, the two basis changes can be replaced by a single one.

We start with the completely observable plant of (8.3-1) and (8.3-2), save that H' is a row vector h'. We do *not* yet assume that $h' = [0 \quad 0 \quad \cdots \quad 0 \quad 1]$ or that the top left-hand $(n-1) \times (n-1)$ submatrix of F has certain desired eigenvalues.

First, with $\det(sI - F) = s^n + a_n s^{n-1} + \cdots + a_1$, define

$$T_1 = \begin{bmatrix} a_2 & a_3 & \cdot & \cdot & a_n & 1 \\ a_3 & a_4 & \cdot & a_n & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & a_n & \cdot & \cdot & \cdot & \cdot \\ a_n & 1 & 0 & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix} \begin{bmatrix} h' \\ h'F \\ h'F^2 \\ \cdot \\ \cdot \\ h'F^{n-1} \end{bmatrix} \tag{8.3-10}$$

This matrix is nonsingular, because $[F, h]$ is completely observable; furthermore, T_1 has the property established in the last section that

$$F_1 = T_1 F T_1^{-1} = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -a_1 \\ 1 & 0 & \cdot & \cdot & 0 & -a_2 \\ 0 & 1 & \cdot & \cdot & 0 & -a_3 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -a_n \end{bmatrix} \qquad h_1 = (T_1^{-1})'h = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}. \tag{8.3-11}$$

We also define $G_1 = T_1 G$, but the precise form of this matrix is of no interest.

Now suppose that the estimator dynamics have been specified, in

terms of the eigenvalues of a matrix F_e, *not yet* defined to be a submatrix of F. These eigenvalues govern the rate at which x_e approaches x. Let the characteristic polynomial of F_e, which is of degree $(n-1)$ and not n, be $\det(sI - F_e) = s^{n-1} + \alpha_{n-1}s^{n-2} + \cdots + \alpha_1$. Define

$$T_2 = \begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 & -\alpha_1 \\ 0 & 1 & \cdot & \cdot & 0 & -\alpha_2 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -\alpha_{n-1} \\ 0 & 0 & \cdot & \cdot & 0 & 1 \end{bmatrix}. \tag{8.3-12}$$

Then it is readily verified that

$$T_2^{-1} = \begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 & \alpha_1 \\ 0 & 1 & \cdot & \cdot & 0 & \alpha_2 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & \alpha_{n-1} \\ 0 & 0 & \cdot & \cdot & 0 & 1 \end{bmatrix} \tag{8.3-13}$$

and

$$F_2 = T_2F_1T_2^{-1} = \left[\begin{array}{ccccc|c} 0 & 0 & \cdot & \cdot & -\alpha_1 & b_1 \\ 1 & 0 & \cdot & \cdot & -\alpha_2 & b_2 \\ 0 & 1 & \cdot & \cdot & -\alpha_3 & b_3 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & b_n \end{array}\right] \tag{8.3-14}$$

where the b_i are entries of a vector b given by

$$b = \begin{bmatrix} 0 \\ \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \alpha_{n-1} \end{bmatrix} - \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \cdot \\ \cdot \\ a_n \end{bmatrix} - \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \cdot \\ \alpha_{n-1} \\ 0 \end{bmatrix}(\alpha_{n-1} - a_n). \tag{8.3-15}$$

Also,

$$h_2' = h_1'T_2^{-1} = [0 \quad 0 \quad \cdot \quad \cdot \quad 0 \quad 1], \tag{8.3-16}$$

which guarantees that y is the last entry of the state vector. Again, we can define $G_2 = T_2G_1$.

The preceding transformations may be interpreted as follows. We started with the plant equations

$$\dot{x} = Fx + Gu \qquad y = h'x. \tag{8.3-17}$$

By setting $x_2 = T_2T_1x$, where T_2T_1 is a nonsingular matrix, we obtained new plant equations

$$\dot{x}_2 = F_2x_2 + G_2u \qquad y = h_2'x_2, \tag{8.3-18}$$

where F_2 and h_2 have the special structures depicted.

Let us temporarily focus now on the problem of estimating x_2. [Clearly, if we have an estimate x_{2e} of x_2, then $x_e = (T_2T_1)^{-1}x_{2e}$ will be an estimate of x, approaching x at the same exponential rate as x_{2e} approaches x_2.] We observe that

$$x_{2e} = \begin{bmatrix} \bar{x}_{2e} \\ y \end{bmatrix} \tag{8.3-19}$$

may be derived from the $(n-1)$-dimensional system

$$\dot{\bar{x}}_{2e} = F_e\bar{x}_{2e} + \bar{b}y + \bar{G}_2u \tag{8.3-20}$$

where F_e is obtained by deleting the last row and column of F_2. That is,

$$F_e = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -\alpha_1 \\ 1 & 0 & \cdot & \cdot & 0 & -\alpha_2 \\ 0 & 1 & \cdot & \cdot & 0 & -\alpha_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -\alpha_{n-1} \end{bmatrix}. \tag{8.3-21}$$

This follows from our earlier discussion; the structure of F_2 given in (8.3-14) and the first of Eqs. (8.3-18) guarantee that $\bar{x}_2$ satisfies the same differential equations as $\bar{x}_{2e}$, and, accordingly, that their difference statisfies

$$\frac{d}{dt}(\bar{x}_2 - \bar{x}_{2e}) = F_e(\bar{x}_2 - \bar{x}_{2e}). \tag{8.3-22}$$

The matrix F_e is the transpose of a companion matrix, and has eigenvalues which are the zeros of $s^{n-1} + \alpha_{n-1}s^{n-2} + \cdots + \alpha_1$. Accordingly, the state estimate x_{2e} converges to x_2 at the desired rate.

To summarize the computations in the preceding scheme, we can list the following steps:

1. Compute the characteristic polynomial of F; form the matrix T_1 of Eq. (8.3-10). Decide on desired estimator dynamics and form the polynomial $s^{n-1} + \alpha_{n-1}s^{n-2} + \cdots + \alpha_1$, whose roots govern the rate at which $x - x_e$ approaches zero. Construct the matrix T_2 of (8.3-12), and transform this coordinate basis, setting $F_2 = T_2T_1FT_1^{-1}T_2^{-1}$, etc.

2. The matrix F_2 will have the form of Eq. (8.3-14). By appropriate partitioning, construct the matrix F_e, consisting of the first $(n - 1)$ rows and columns of F, and the vector $\bar{b}$, the last column of F except for the last element. Define $\bar{G}_2$ as G_2 less its last row.
3. Use F_e, $\bar{b}$, and $\bar{G}_2$ to implement Eq. (8.3-20). Obtain x_{2e} as

$$\begin{bmatrix} \bar{x}_{2e} \\ y \end{bmatrix}$$

and x_e as $(T_2T_1)^{-1}x_{2e}$.

Figure 8.3-2 shows the detailed estimator structure.

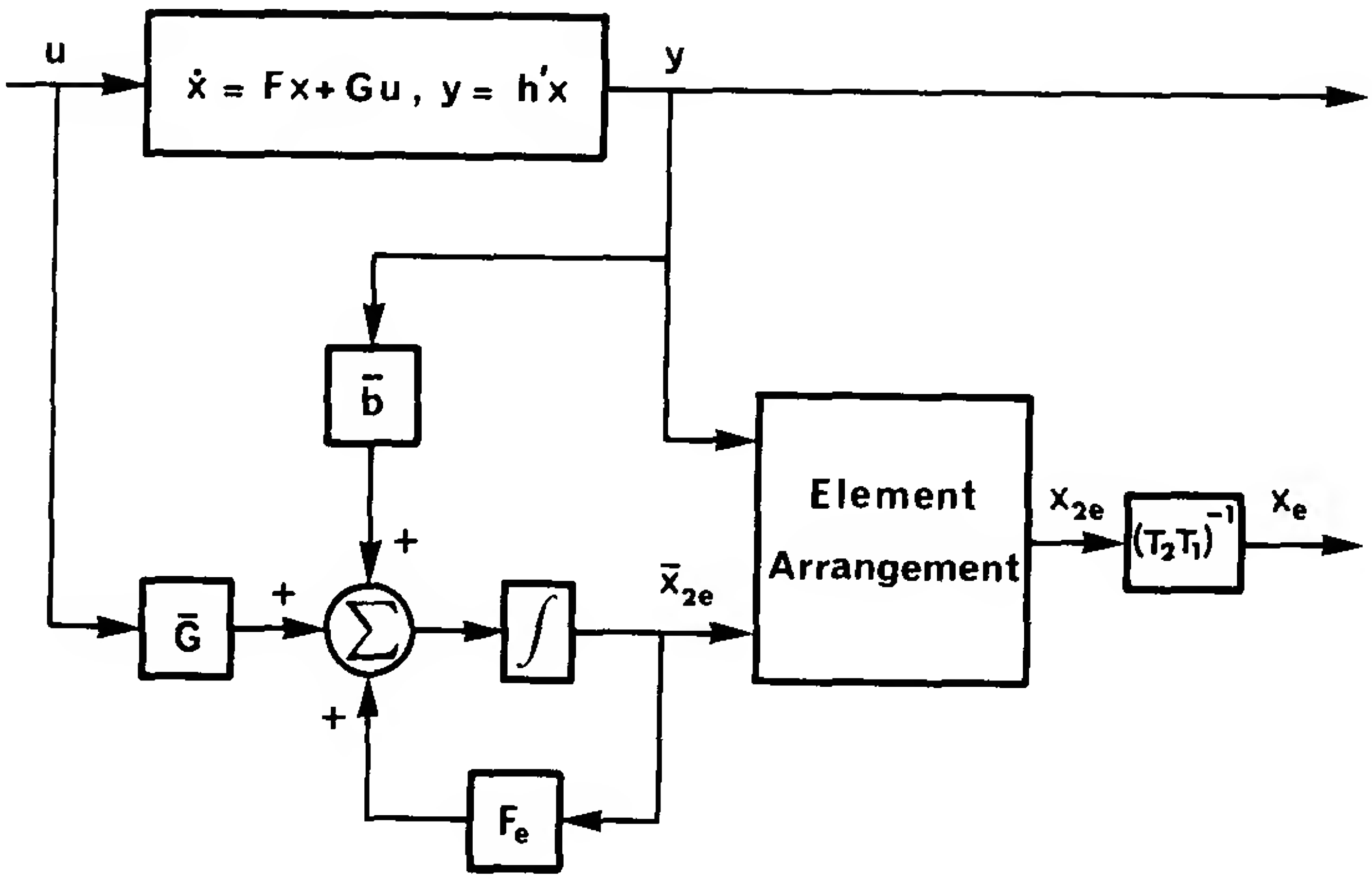


Fig. 8.3-2 Detailed structure of the single-output Luenberger estimator.

Let us now consider a simple example. Suppose we are given the plant equations

$$\dot{x} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u$$

$$y = [2 \quad -1]x.$$

In the last section, we found that with

$$T_1 = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix},$$

we had

$$F_1 = T_1FT_1^{-1} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \qquad h_1 = (T_1^{-1})'h = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Also,

$$g_1 = T_1 g = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

There is only one eigenvalue to specify for the matrix F_e. Suppose we take it at -10. Then

$$T_2 = \begin{bmatrix} 1 & -10 \\ 0 & 1 \end{bmatrix},$$

which leads to

$$T_2^{-1} = \begin{bmatrix} 1 & +10 \\ 0 & 1 \end{bmatrix}, \quad F_e = T_2 F_1 T_2^{-1} = \begin{bmatrix} -10 & -110 \\ 1 & 11 \end{bmatrix},$$

$$h_2 = (T_2^{-1})' h_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and

$$g_2 = T_2 g_1 = \begin{bmatrix} -9 \\ 1 \end{bmatrix}.$$

From (8.3-14), the equation for $\bar{x}_{2e}$ is

$$\dot{\bar{x}}_{2e} = -10\bar{x}_{2e} - 110y - 9u$$

and

$$x_e = (T_2 T_1)^{-1} x_{2e} = \begin{bmatrix} \frac{1}{2} & \frac{11}{2} \\ 1 & 10 \end{bmatrix} \begin{bmatrix} x_{2e} \\ y \end{bmatrix}.$$

Note that the calculation of F_1, g_1, and h_1 is not, in fact, necessary. It is sufficient to find T_1 and T_2, and proceed straight to F_2, etc., via $F_2 = T_2 T_1 F T_1^{-1} T_2^{-1}$. The plant and estimator are shown in Fig. 8.3-3.

We shall now give an incomplete discussion of estimators for multiple-output plants. A complete discussion can be found in [9]. Even if the plant states are observable from one of a number of outputs, it is preferable not to employ an observer using just this one output, on the grounds that if more output components are used, the dimension of the estimator can be lowered. In general, if there are r independent outputs, the estimator need only be of dimension $n - r$, and the greater the number of outputs available, the less the dimension of the estimator.

The case of a plant composed of a cascade of smaller plants, with the output of each smaller plant available, is easy to consider using the theory for single-output plants. Thus, consider the arrangement of Fig. 8.3-4, where n_i is the dimension of the small plant i. Estimation of the overall plant state may be broken up into estimation of the states of a number of scalar output plants, the estimator for the small plant i being of dimension $n_i - 1$. Consequently, the overall estimator dimension is $\sum (n_i) - r$—i.e., r less than the

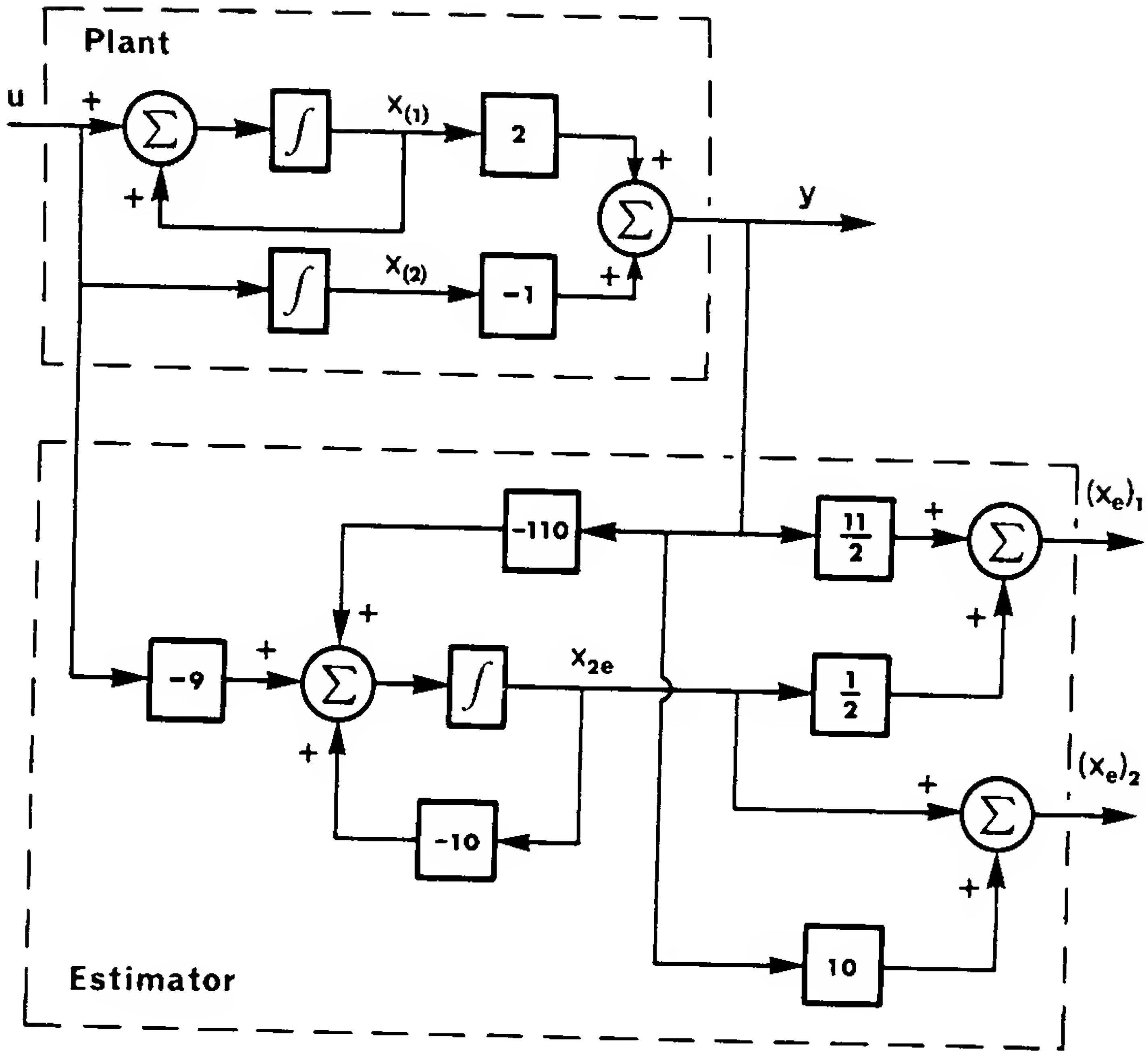


Fig. 8.3-3 Second order plant example with estimator.

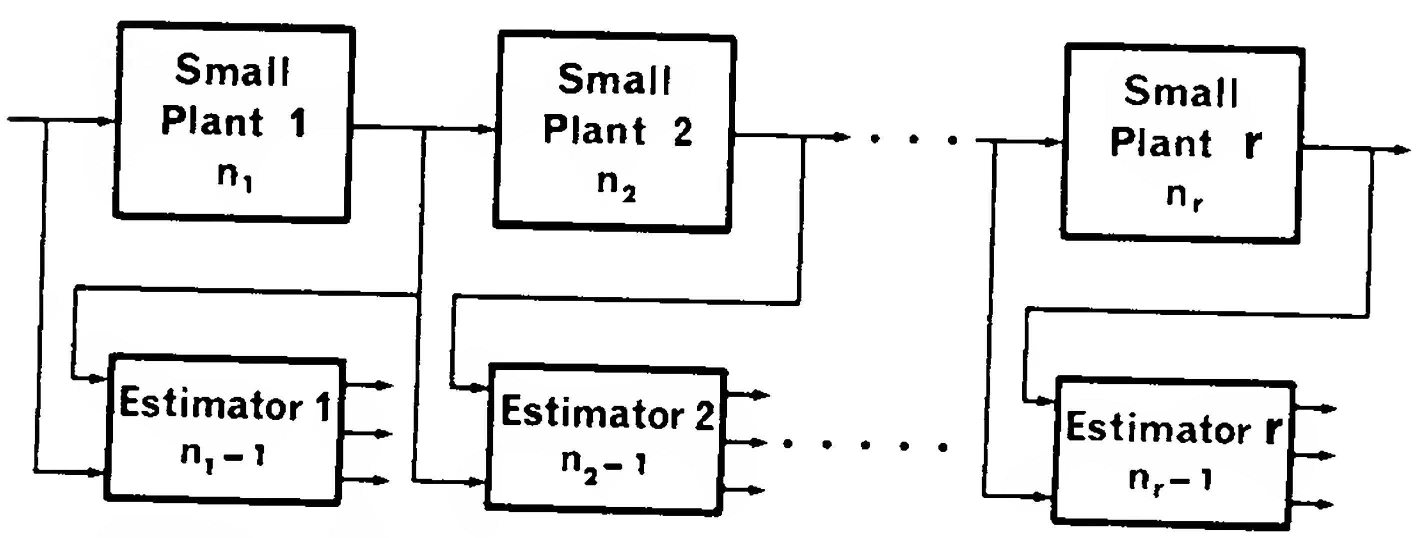


Fig. 8.3-4 State estimation for a special multiple-output plant.

dimension of the overall plant—which may represent a substantial saving in complexity.

All the reasons listed in the last section in favor of estimating this way, rather than using merely the overall plant input and output, are applicable here. In addition, the complexity of the estimator in this case is simpler. In the next chapter, we shall discuss how further simplification again of the estimator structure may be achieved when all that is required is a linear function $k'x_e$ of the components of the state estimator, rather than all components separately.

We now give qualitative comments on the effect of noise in the Luen-

berger estimator. As for the estimators of the last section, it will be noise that limits the extent to which the eigenvalues of F_e can be made negative, i.e., the rapidity with which x_e will approach x. But there is a second factor present in the estimator of this section which is absent in those of the last section. Suppose the plant output y includes white noise. Using the scheme of the last section, the resulting noise in x_e is bandlimited, essentially because there is always integration between y and x_e. This is not the case with the estimators of this section: To be sure, there will be bandlimited noise in x_e arising from white noise in y, but there will also be white noise in x_e arising from that in y, because x_e is partly determined by a memoryless transformation on y. Consequently, one might expect the performance of the estimators of this section to be worse than those of the previous section in a noisy environment.

Problem 8.3-1: Consider the scheme of Fig. 8.3-5. Write down state-space equations for the individual blocks, and design a Luenberger state estimator assuming availability of the signals y_1 and y_2.

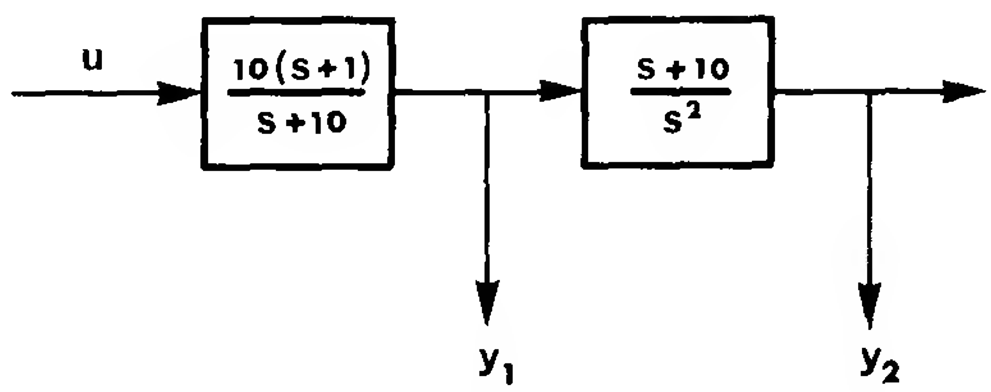


Fig. 8.3-5 Plant for Problem 8.3-1.

Problem 8.3-2: Design a Luenberger estimator for

$$\dot{x} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix} x + \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} u$$

$$y = [1 \quad 1 \quad 2]x.$$

Problem 8.3-3: Consider the plant

$$\dot{x} = \begin{bmatrix} -2 & 1 \\ 0 & -2 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

$$y = [1 \quad 0]x$$

and assume that associated with y is white noise, i.e., noise with power spectral density $S(\omega) = \sigma$. Design a Luenberger estimator with the F_e matrix equal to $-\alpha$ for a positive constant α. Compute the spectral density of noise in both components of x_e as a function of α. [*Hint:* Obtain the transfer functions relating y to each component of x_e. If these are $t_1(j\omega)$ and $t_2(j\omega)$, the spectral density of the noise in the components is $\sigma|t_1(j\omega)|^2$ and $\sigma|t_2(j\omega)|^2$.]

Problem 8.3-4: Repeat Problem 8.3-3, with the replacement of the Luenberger estimator by a second-order estimator with $F + k_e h'$ possessing two eigenvalues at $-\alpha$. Compare the resulting noise densities with those of Problem 8.3-3.

8.4 STATISTICAL ESTIMATOR DESIGN (THE KALMAN-BUCY FILTER)

In this section, we touch upon an enormous body of knowledge perhaps best described by the term "filtering theory." Much of this theory is summarized in the two books [11] and [12]. The particular material covered here is discussed in the important paper [13], although to carry out certain computations, we make use of a method discussed in [14] and [15].

Broadly speaking, we shall attempt to take quantitative consideration of the noise associated with measurements on a plant when designing an estimator. This means that the design of the estimator depends on probabilistic data concerning the noise. We shall also aim at building the best possible estimator, where by "best possible" we mean roughly the estimator whose output is closest to what it should be, despite the noise. In other words, we shall be attempting to solve an *optimal filtering*, as distinct from optimal control, problem.

We warn the reader in advance of two things:

1. The treatment we shall give will omit many insights, side remarks, etc., in the interests of confining the discussion to a reasonable length.
2. The discussion will omit some details of mathematical rigor. We shall perform integration operations with integrands involving random variables, and the various operations, although certainly valid for deterministic variables, need to be proved to be valid for random variables. However, we shall omit these proofs. Moreover, we shall interchange integration and expectation operations without verifying that the interchanges are permissible.

In outline, we shall first describe the optimal estimation or filtering problem—i.e., we shall describe the systems considered, the associated noise statistics, and a specific estimation task. Then, by the introduction of new variables, we shall convert the filtering problem into a deterministic optimal regulator problem, of the sort we have been discussing all through this book. The solution of this regulator problem will then yield a mathematical solution of the filtering problem. A technique for physical implementation of the solution will then be found, leading to an estimator structure of the same form as that considered in Sec. 8.2, except that noise is present at certain points in the plant, and as a consequence in the estimator, see Fig. 8.4-1. (The figure

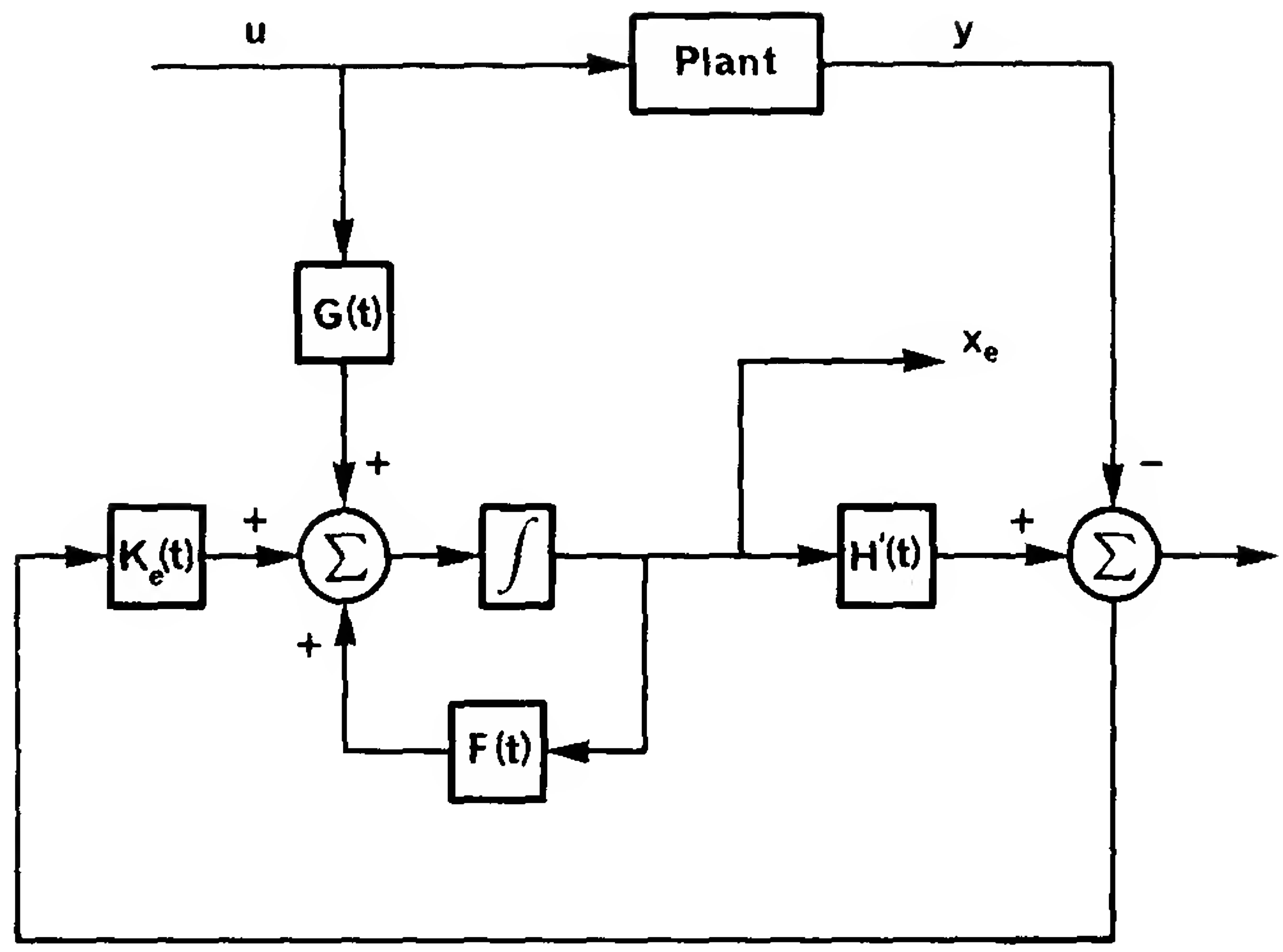


Fig. 8.4-1 Structure of optimal estimator.

does not show the structure of the plant, which is assumed to be of the standard form $\dot{x} = Fx + Gu$, $y = H'x$, with additional terms in these equations representing noise, to be indicated precisely later.)

Since the structures of the optimal estimator and the estimator of Sec. 8.2 are the same, we can regard the present section as describing a technique for *optimally* designing one of the estimators of Sec. 8.2. The computations required for optimal design are a good deal more involved than those for the earlier design for a time-invariant single-output plant. However, for a multiple-output plant it is possible that the calculations to be presented here might even be simpler than the appropriate multiple-output plant generalization of the calculations of Sec. 8.2. The calculations here also extend to time-varying plants.

Description of plants and noise statistics. The plants we shall consider are of the form

$$\frac{dx(t)}{dt} = F(t)x(t) + G(t)u(t) + v(t) \tag{8.4-1}$$

$$y(t) = H'(t)x(t) + w(t). \tag{8.4-2}$$

Here, $v(t)$ and $w(t)$ represent noise terms, which will be explained shortly. The dependence of F, G, and H on t, indicated in the equations, is to emphasize that, at least for the moment, these quantities are not necessarily time invariant. However, for infinite-time interval problems, which are considered later, we shall specialize to the time-invariant case. Without further comment, we assume that the entries of $F(\cdot)$, $G(\cdot)$, and $H(\cdot)$ are all continuous. There

is no restriction on the dimensions of u and y in these equations, and the subsequent calculations will not, in fact, be simplified significantly by an assumption that either u or y is scalar.

The properties of the noise terms will now be discussed. First note that the model of (8.4-1) and (8.4-2) assumes additive noise only, and it also assumes that noise is injected at only two points (see Fig. 8.4-2). The latter restriction is not so severe as might at first appear. Thus, e.g., any noise entering with $u(t)$ [and passing through the $G(t)$ block] is equivalent to some other noise entering at the same point as $v(t)$.

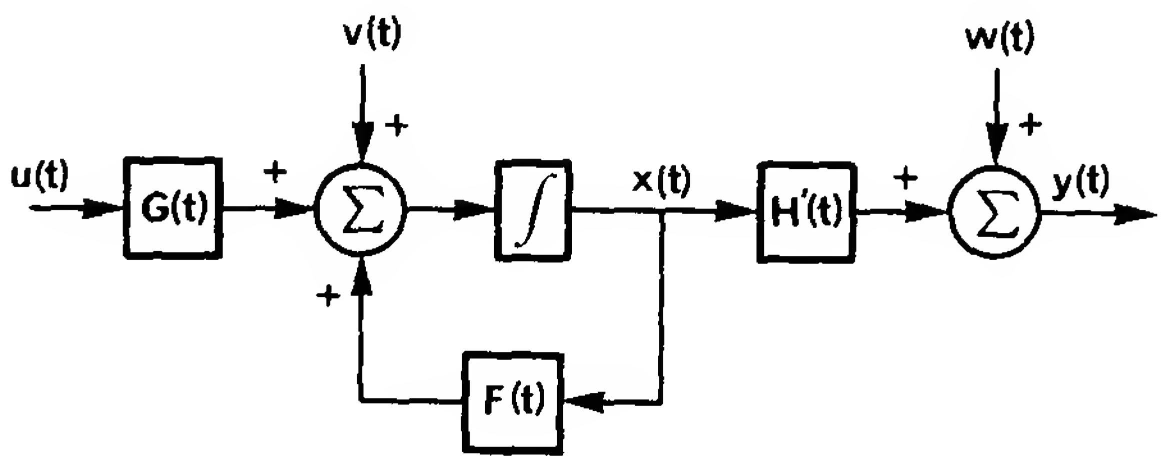


Fig. 8.4-2 Plant with additive noise.

In the case of both $v(t)$ and $w(t)$, the noise is assumed to be white, gaussian, and to have zero mean. The first property implies that it is uncorrelated from instant to instant; if it were also stationary, it would have a constant power spectrum. The second property implies that all probabilistic information about the noise is summed up in the covariance of the noise—viz., $E[v(t)v'(\tau)]$ for $v(t)$ and likewise for $w(t)$. This convenient mathematical assumption is fortunately not without physical basis, for many naturally occurring noise processes are, indeed, gaussian, and such processes normally have zero mean. Therefore, in mathematical terms,

$$E[v(t)v'(\tau)] = Q(t)\delta(t-\tau) \quad E[v(t)] \equiv 0 \tag{8.4-3}$$

$$E[w(t)w'(\tau)] = R(t)\delta(t-\tau) \quad E[w(t)] \equiv 0 \tag{8.4-4}$$

for some matrices $Q(\cdot)$ and $R(\cdot)$, which we assume without further comment to have all entries continuous. The presence of the $\delta(t-\tau)$ term guarantees the whiteness property. Precisely because the quantities on the left sides of (8.4-3) and (8.4-4) are *covariances*, the matrices $Q(t)$ and $R(t)$ must be symmetric and nonnegative definite. But we shall make the additional assumption that $R(t)$ is positive definite for all t. If this were not the case, there would be some linear combination of the outputs that was entirely noise free. Then, in an appropriate coordinate basis, one entry of the state vector would be known without filtering. As a result, the optimal estimator would not have the structure of Fig. 8.4-1 and a different optimal estimator design procedure

would be required. We shall omit consideration of this difficult problem here; the interested reader should consult reference [16].

Because it is often the case physically, we shall assume that the noise processes $v(t)$ and $w(t)$ are independent. This means that

$$E[v(t)w'(\tau)] = 0 \qquad \text{for all } t \text{ and } \tau. \tag{8.4-5}$$

The final assumptions concern the initial state of (8.4-1). State estimation is assumed to commence at some time t_0, which may be minus infinity or may be finite. It is necessary to assume something about the state $x(t_0)$, and the assumptions that prove of use are that $x(t_0)$ is a gaussian random variable, of mean m, and covariance P_0—i.e.,

$$E\{[x(t_0) - m][x(t_0) - m]'\} = P_0 \qquad E[x(t_0)] = m. \tag{8.4-6}$$

Furthermore, $x(t_0)$ is independent of $v(t)$ and $w(t)$—i.e.,

$$E[x(t_0)v'(t)] = E[x(t_0)w'(t)] = 0 \qquad \text{for all } t. \tag{8.4-7}$$

Notice that the case where $x(t_0)$ has a known (deterministic) value is included: If $P_0 = 0$, then $x(t_0) = m$, rather than just $E[x(t_0)] = m$.

Let us now summarize the plant and noise descriptions.

ASSUMPTION 8.4-1.

1. The plant is described by the equations (8.4-1) and (8.4-2).
2. The noise processes $v(t)$ and $w(t)$ are white, gaussian, of zero mean, are independent, and have known covariances [see Eqs. (8.4-3), (8.4-4), and (8.4-5)]. The matrix $R(t)$ in (8.4-4) is nonsingular.
3. The initial state of the plant is a gaussian random variable, of known mean and covariance [Eq. (8.4-6)]. It is independent of $v(t)$ and $w(t)$ [see Eq. (8.4-7)].

Notice that an assumption of complete observability of the plant has not been made. Such will be made later in considering infinite-time problems.

Although all the assumptions made often have some physical validity, there will undoubtedly be many occasions when this is not the case. Many associated extensions of the preceding problem formulation have, in fact, been considered, and optimal estimators derived, but to consider these would be to go well beyond the scope of this book.

Statement of the optimal estimation problem. We shall now define the precise task of estimation. The information at our disposal consists of the plant input $u(t)$ and output $y(t)$ for $t_0 \leq t \leq t_1$, and the probabilistic descriptions of $x(t_0)$, $v(t)$, and $w(t)$. To obtain a first solution to the estimation problem, it proves convenient to make two temporary simplifications.

TEMPORARY ASSUMPTION 8.4-2. The external input to the plant $u(t)$ is identically zero, and the mean m of the initial state $x(t_0)$ is zero.

and

TEMPORARY ASSUMPTION 8.4-3. The initial time t_0 is finite.

These assumptions will be removed when we have derived the optimal estimator for the special case implicit in the assumptions. With Temporary Assumptions 8.4-2 and 8.4-3 in force, the data at our disposal is simply the plant output $y(t)$ for $t_0 \leq t \leq t_1$, and our knowledge of the covariances of $v(\cdot)$, $w(\cdot)$, and $x(t_0)$.

Now, since $x(t_1)$ is a vector, it is convenient to regard the problem of its estimation as a problem of estimating separately each of its components. Therefore, we shall consider the estimation of $b'x(t_1)$, where b is an arbitrary constant vector. Special choices of b, of course, lead to estimates of the various components of $x(t_1)$. The sort of estimate for which we shall aim is a minimum variance estimate—i.e., we want to construct from a measurement of $y(t)$, $t_0 \leq t \leq t_1$, a certain number, call it β, such that

$$E[b'x(t_1) - \beta]^2$$

is minimum. Then β is the minimum variance estimate of $b'x(t_1)$. It turns out that because all the random processes and variables are gaussian, and have zero mean, the number β can be derived by linear operations on $y(t)$, $t_0 \leq t \leq t_1$—i.e., there exists some function $s(t; b, t_1)$, $t_0 \leq t \leq t_1$ such that

$$\beta = \int_{t_0}^{t_1} s'(t; b, t_1)y(t)\, dt.$$

(This is a deep result which we shall not prove here.) The introduction of $s(\cdot; b, t_1)$ now allows the following formal statement of the optimal estimation problem.

> ***Optimal estimation problem.*** Given the plant of Eqs. (8.4-1) and (8.4-2), suppose that Assumptions 8.4-1, 8.4-2, and 8.4-3 hold. Then, for fixed but arbitrary values of b and $t_1 \geq t_0 > -\infty$, find a vector function of time $s(t; b, t_1)$, $t_0 \leq t \leq t_1$, of the same dimension as $y(t)$, such that
>
> $$E\left\{\left[b'x(t_1) - \int_{t_0}^{t_1} s'(t; b, t_1)y(t)\, dt\right]^2\right\}$$
>
> is minimized, the expectation being over all possible realizations of the two noise processes and over the random variable $x(t_0)$. [A minimum variance estimate of $b'x(t_1)$ is then provided by $\int_{t_0}^{t_1} s'(t; b, t_1)y(t)\, dt$].

A further problem is to state how this estimate might be physically implemented to produce an on-line estimate $x_e(t_1)$ at time t_1 of $x(t_1)$, which is continuously updated, rather than a single estimate of the scalar random variable $b'x(t_1)$ for fixed b and t_1.

Without further comment, we shall use the notation $s(t)$ and $s(\cdot)$ as shorthand for $s(t; b, t_1)$ and $s(\cdot\,; b, t_1)$, provided no ambiguity occurs.

At this stage, let us pause to review what we have done so far, and what we shall do in the next part of this section. Thus far, we have accomplished the following:

1. We have described the plants considered, together with the noise associated with the plants.
2. We have posed a problem of estimating a *particular* linear combination of the entries of the state vector at a *particular* time [i.e., $b'x(t_1)$ for fixed b and t_1], using input and output measurements till t_1. The estimate is to be a minimum variance one.
3. We have posed the problem of constructing a device which at every time t_1 will produce an estimate of $x(t_1)$, rather than $b'x(t_1)$. That is, we have posed the problem of constructing an on-line estimate of $x(t_1)$.

In the remainder of this section, we shall do the following:

1. We shall show how the first of the preceding optimal estimation problems can be reformulated as an optimal control problem. *We caution the reader that the existence of such a reformulation is probably not intuitively reasonable, and the parameters appearing in the regulator problem are only suggested by hindsight.* Therefore, the reader will have to suppress such natural questions as "Why pick such-and-such set of system equations?" and remind himself that the justification for the choice of such-and-such system equation lies in the fact that it works, somewhat surprisingly.
2. Using the optimal regulator problem reformulation, and our knowledge of the general regulator problem solution, we shall solve the specific optimal regulator problem.
3. Next, we shall do a natural thing—use the solution of the regulator problem to write down a solution of the optimal estimation problem associated with estimating $b'x(t_1)$ for specific b and t_1.
4. We shall then show how to obtain an estimate of $x(t_1)$ on line.
5. Elimination of restrictive assumptions, examples, and some extensions will then follow.

Reformulation of the optimal estimation problem as an optimal regulator problem. To carry through the reformulation, we introduce a new vector function of time $r(\cdot)$, of the same dimension as $x(\cdot)$. This function is defined from $s(\cdot)$ via the equation

$$\frac{d}{dt}r(t) = -F'(t)r(t) + H(t)s(t) \qquad r(t_1) = b. \tag{8.4-8}$$

Observe that Eq. (8.4-8) has the same form as the equation $\dot{x} = Fx + Gu$,

with prescribed boundary condition $x(t_0)$, except that we shall be interested in the solutions of (8.4-8) for $t \leq t_1$ rather than for $t \geq t_1$.

We shall now rewrite $E[b'x(t_1) - \int_{t_0}^{t_1} s'(t)y(t)\,dt]^2$ as a quadratic performance index involving $r(\cdot)$ and $s(\cdot)$. From (8.4-1) and (8.4-8), we have

$$\begin{aligned}\frac{d}{dt}[r'(t)x(t)] &= \dot{r}'(t)x(t) + r'(t)\dot{x}(t)\\ &= -r'Fx + s'H'x + r'Fx + r'v\\ &= s'y - s'w + r'v.\end{aligned}$$

Integrating this equation from t_0 to t_1, and using the boundary condition on r, leads to

$$b'x(t_1) - r'(t_0)x(t_0) = \int_{t_0}^{t_1} s'(t)y(t)\,dt - \int_{t_0}^{t_1} s'(t)w(t)\,dt + \int_{t_0}^{t_1} r'(t)v(t)\,dt$$

or

$$b'x(t_1) - \int_{t_0}^{t_1} s'(t)y(t)\,dt = r'(t_0)x(t_0) - \int_{t_0}^{t_1} s'(t)w(t)\,dt + \int_{t_0}^{t_1} r'(t)v(t)\,dt. \tag{8.4-9}$$

The next step is to square each side of (8.4-9) and take the expectation. Because of the independence of $x(t_0)$, $w(t)$, and $v(t)$, there results

$$\begin{aligned}E\left[b'x(t_1) - \int_{t_0}^{t_1} s'(t)y(t)\,dt\right]^2 &= E\{[r'(t_0)x(t_0)]^2\}\\ &\quad + E\left\{\left[\int_{t_0}^{t_1} s'(t)w(t)\,dt\right]^2\right\}\\ &\quad + E\left\{\left[\int_{t_0}^{t_1} r'(t)v(t)\,dt\right]^2\right\}.\end{aligned}$$

The three expectations on the right side of this equation are easy to evaluate, assuming interchangeability of the expectation and integration operation in the case of the second and third. Thus,

$$\begin{aligned}E\{[r'(t_0)x(t_0)]^2\} &= E[r'(t_0)x(t_0)x'(t_0)r(t_0)]\\ &= r'(t_0)E[x(t_0)x'(t_0)]r(t_0)\\ &= r'(t_0)P_0r(t_0).\end{aligned}$$

Also,

$$\begin{aligned}E\left\{\left[\int_{t_0}^{t_1} s'(t)w(t)\,dt\right]^2\right\} &= E\left[\int_{t_0}^{t_1}\int_{t_0}^{t_1} s'(t)w(t)w'(\tau)s(\tau)\,dt\,d\tau\right]\\ &= \int_{t_0}^{t_1}\int_{t_0}^{t_1} s'(t)E[w(t)w'(\tau)]s(\tau)\,dt\,d\tau\\ &= \int_{t_0}^{t_1}\int_{t_0}^{t_1} s'(t)R(t)\delta(t-\tau)s(\tau)\,dt\,d\tau\\ &= \int_{t_0}^{t_1} s'(t)R(t)s(t)\,dt.\end{aligned}$$

Similarly,

$$E\left\{\left[\int_{t_0}^{t_1} r'(t)v(t)\,dt\right]^2\right\} = \int_{t_0}^{t_1} r'(t)Q(t)r(t)\,dt.$$

Therefore,

$$E\left\{\left[b'x(t_1) - \int_{t_0}^{t_1} s'(t)y(t)\,dt\right]^2\right\}$$
$$= r'(t_0)P_0r(t_0) + \int_{t_0}^{t_1} [s'(t)R(t)s(t) + r'(t)Q(t)r(t)]\,dt. \tag{8.4-10}$$

Now all quantities on the right of (8.4-10) are deterministic, with $s(\cdot)$ free to be chosen and $r(\cdot)$ related to $s(\cdot)$ via (8.4-8). Therefore, the problem of choosing $s(\cdot)$ to minimize the left side of (8.4-10) is the same as the *deterministic* problem of choosing $s(\cdot)$ to minimize the right side of (8.4-10), subject to (8.4-8) holding.

Let us summarize the reformulation of the optimal estimation problem as follows.

> ***Reformulated optimal estimation problem.*** Suppose we are given the plant of Eqs. (8.4-1) and (8.4-2), and suppose that Assumptions 8.4-1, 8.4-2, and 8.4-3 hold. Let b be an arbitrary constant vector and $t_1 > t_0$ be an arbitrary time. Find a function $s(t; b, t_1)$ for $t_0 \leq t \leq t_1$ such that the (deterministic) performance index (8.4-10) is minimized, subject to Eq. (8.4-8) holding.

We shall now comment upon this reformulated problem. Recall that R is positive definite for all t; Q is nonnegative definite for all t; and P_0, being the covariance of a vector random variable, is also nonnegative definite. Therefore, the only factors distinguishing the problem of finding the function $s(\cdot)$, which minimizes (8.4-10) subject to (8.4-8), from the usual optimal regulator problem are (1) that the boundary condition on $r(t)$ occurs at the final time t_1 rather than the initial time t_0, and (2) that the term $r'(t_0)P_0r(t_0)$ represents an initial rather than a final value term. In a sense, the problem we face here is a regulator problem with time running backward. We shall solve the minimization problem by reversing the time variable with a suitable transformation. This will give a solution to the minimization problem in feedback form, i.e., $s(t)$ will be expressed as a certain (linear) function of $r(t)$. Then, use of the differential equation for $r(t)$ will allow us to write down an explicit formula for $s(t)$.

Solution to the regulator problem in feedback form. We shall now compute the solution of the minimization problem associated with (8.4-8) and (8.4-10). Define a new running variable $\hat{t} = -t$, and vectors $\hat{s}(\hat{t})$, $\hat{r}(\hat{t})$, and matrices $\hat{F}(\hat{t})$, $\hat{R}(\hat{t})$, etc., by $\hat{s}(\hat{t}) = s(t)$, $\hat{r}(\hat{t}) = r(t)$, and $\hat{F}(\hat{t}) = F(t)$,

$\hat{R}(\hat{t}) = R(t)$, etc., when $\hat{t} = -t$. The performance index to be minimized is

$$V = \hat{r}'(\hat{t}_0)P_0\hat{r}(\hat{t}_0) + \int_{\hat{t}_1}^{\hat{t}_0} [\hat{s}'(\hat{t})\hat{R}(\hat{t})\hat{s}(\hat{t}) + \hat{r}'(\hat{t})\hat{Q}(\hat{t})\hat{r}(\hat{t})]\, d\hat{t} \qquad (8.4\text{-}11)$$

where $\hat{t}_1 < \hat{t}_0$, and the constraint equations are

$$\frac{d}{d\hat{t}}\hat{r}(\hat{t}) = \hat{F}'(\hat{t})\hat{r}(\hat{t}) - \hat{H}(\hat{t})\hat{s}(\hat{t}) \quad \hat{r}(\hat{t}_1) = b. \qquad (8.4\text{-}12)$$

Minimization of (8.4-11) subject to (8.4-12) is achieved as follows. Let $\hat{P}(\hat{t})$ be the solution of

$$-\frac{d}{d\hat{t}}\hat{P} = \hat{P}\hat{F}' + \hat{F}\hat{P} - \hat{P}\hat{H}\hat{R}^{-1}\hat{H}'\hat{P} + \hat{Q} \quad \hat{P}(\hat{t}_0) = P_0 \qquad (8.4\text{-}13)$$

where the argument $\hat{t}$ of the matrices in (8.4-13) is suppressed. Then the control law

$$\hat{s}(\hat{t}) = \hat{R}^{-1}(\hat{t})\hat{H}'(\hat{t})\hat{P}(\hat{t})\hat{r}(\hat{t}) \quad \hat{t}_1 \leq \hat{t} \leq \hat{t}_0 \qquad (8.4\text{-}14)$$

achieves the minimization. Because $\hat{R}(\hat{t})$ is positive definite, and $\hat{Q}(\hat{t})$ and P_0 are nonnegative definite, it follows from the earlier regulator theory that $\hat{P}(\hat{t})$ exists for all $\hat{t} \leq \hat{t}_0$ and is symmetric nonnegative definite. Now, with the definition $P(t) = \hat{P}(\hat{t})$ when $\hat{t} = -t$, Eq. (8.4-13) is readily shown to imply

$$\frac{dP}{dt} = PF' + FP - PHR^{-1}H'P + Q \quad P(t_0) = P_0 \qquad (8.4\text{-}15)$$

where the suppressed argument is now t. Consequently, (8.4-14) implies

$$s(t) = R^{-1}(t)H'(t)P(t)r(t) \qquad (8.4\text{-}16)$$

and we conclude that the function $s(\cdot)$ minimizing (8.4-10), subject to (8.4-8), satisfies Eq. (8.4-16), with P defined by (8.4-15). We observe, too, that the existence of $P(t)$ as the solution of (8.4-15) is guaranteed for all $t \geq t_0$, because the existence of $\hat{P}(\hat{t})$ is guaranteed for all $\hat{t} \leq \hat{t}_0$; $P(t)$, of course, is also symmetric nonnegative definite.

Explicit solution of the optimal estimation problem. To complete the solution of the optimal estimator problem, we must evaluate the function $s(t)$ as an explicit function of time. Inserting (8.4-16) into (8.4-8), we have

$$\frac{d}{dt}r(t) = -[F(t) - P(t)H(t)R^{-1}(t)H'(t)]'r(t) \quad r(t_1) = b. \qquad (8.4\text{-}17)$$

Let us define

$$F_e(t) = F(t) - P(t)H(t)R^{-1}(t)H'(t) \qquad (8.4\text{-}18)$$

and let $\Phi_e(t, \tau)$ be the transition matrix associated with $\dot{x} = F_e x$. Then the solution of (8.4-17) can readily be verified to be

$$r(t) = \Phi_e'(t_1, t)b.$$

Thus, from (8.4-16),

$$s(t) = R^{-1}(t)H'(t)P(t)\Phi_e'(t_1, t)b. \tag{8.4-19}$$

The manipulations between Eqs. (8.4-16) and (8.4-19) serve two purposes: (1) to introduce the matrix $F_e(t)$; and (2) to express the function $s(\cdot)$ minimizing (8.4-10) subject to (8.4-8) as an explicit function of time, rather than as a feedback law.

Since the solutions to the optimal estimation problem and reformulated optimal estimation problem are the same, we shall merely summarize the solution to the former.

Solution to the optimal estimation problem. Given the plant equations (8.4-1) and (8.4-2), suppose that Assumptions 8.4-1, 8.4-2, and 8.4-3 hold. Then for fixed but arbitrary values of b and $t_1 \geq t_0$, the function of time $s(t; b, t_1)$, $t_0 \leq t \leq t_1$, which minimizes

$$E\left\{\left[b'x(t_1) - \int_{t_0}^{t_1} s'(t; b, t_1)y(t)\,dt\right]^2\right\},$$

is defined as follows. Let $P(t)$ be the solution of the Riccati equation (8.4-15), which exists and is symmetric nonnegative definite for all $t \geq t_0$. Let $F_e(t)$ be defined via (8.4-18), and let $\Phi_e(t, \tau)$ be the transition matrix associated with $F_e(t)$. Then $s(t; b, t_1)$ is given by Eq. (8.4-19), repeated for convenience:

$$s(t; b, t_1) = R^{-1}(t)H'(t)P(t)\Phi_e'(t_1, t)b. \tag{8.4-19}$$

Construction of an on-line optimal estimate. We shall now examine the problem of physically implementing the estimation procedure on line. Since the minimum variance estimate of $b'x(t_1)$ is $b' \int_{t_0}^{t_1} \Phi_e(t_1, t)P(t)H(t)R^{-1}(t)y(t)\,dt$ for *arbitrary* b, it follows by taking $b' = e_i'$ (a row vector consisting of zeros, save for unity in the ith position) that the minimum variance estimate of the ith entry of $x(t_1)$ is the ith entry of $\int_{t_0}^{t_1} \Phi_e(t_1, t)P(t)H(t)R^{-1}(t)y(t)\,dt$. In other words,

$$x_e(t_1) = \int_{t_0}^{t_1} \Phi_e(t_1, t)P(t)H(t)R^{-1}(t)y(t)\,dt. \tag{8.4-20}$$

On-line implementation of this estimate is now simple, for observe that (8.4-20) is equivalent to requiring

$$\frac{d}{dt}x_e(t) = F_e(t)x_e(t) - K_e(t)y(t) \quad x_e(t_0) = 0 \tag{8.4-21}$$

where

$$K_e(t) = -P(t)H(t)R^{-1}(t). \tag{8.4-22}$$

Figure 8.4-3(a) shows a realization of this equation, which we reiterate is valid, provided that $u(t) \equiv 0$, $E[x(t_0)] = m = 0$, and t_0 is finite. If we observe

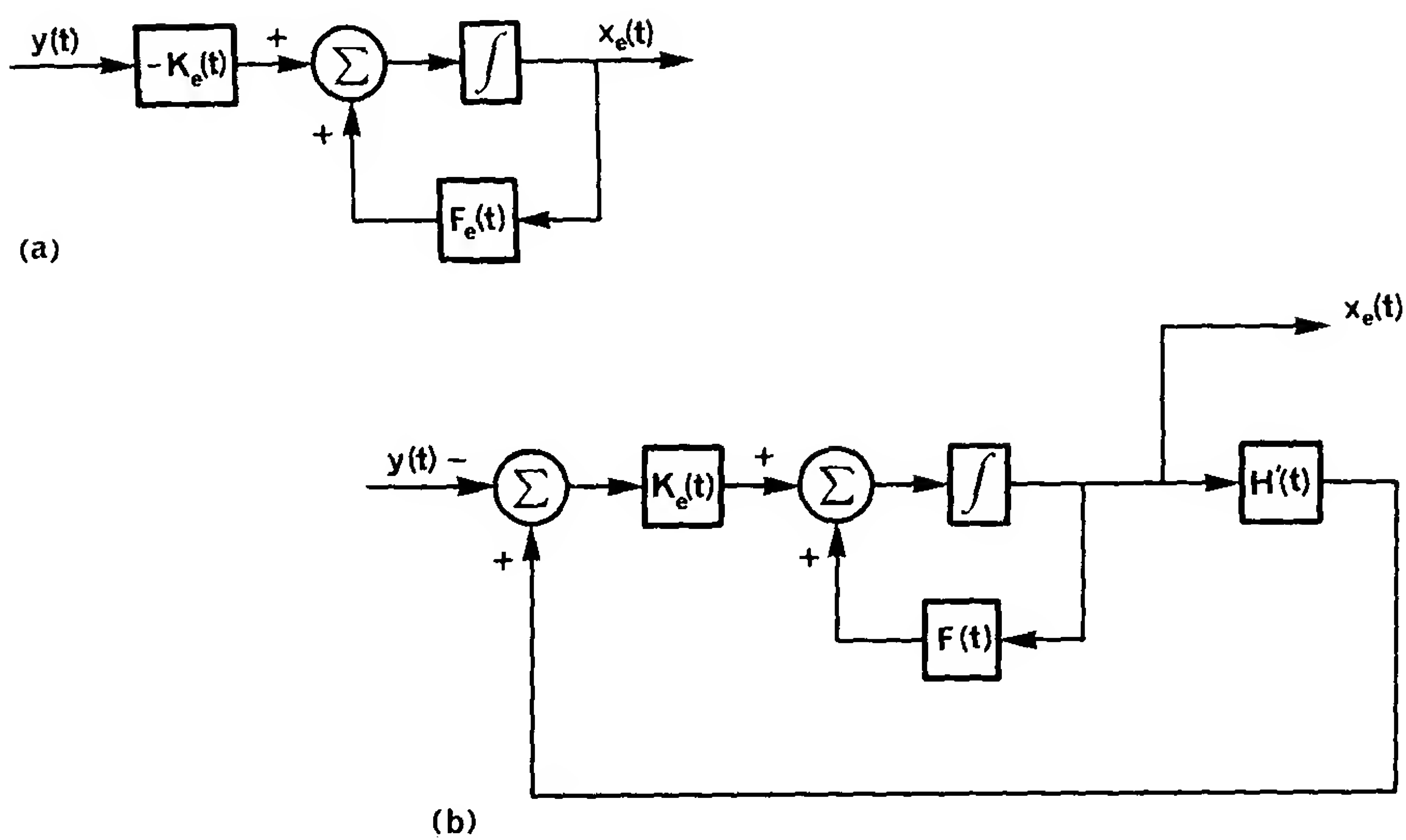


Fig. 8.4-3 (a) First estimator structure [$u(t) \equiv 0,\ m = 0$]; (b) second estimator structure [$u(t) \equiv 0,\ m = 0$].

from (8.4-18) that $F_e(t) = F(t) + K_e(t)H'(t)$, it follows that (8.4-21) is equivalent to

$$\frac{d}{dt}x_e(t) = F(t)x_e(t) + K_e(t)[H'(t)x_e(t) - y(t)] \quad x_e(t_0) = 0. \tag{8.4-23}$$

Figure 8.4-3(b) shows the associated estimator realization.

We have now covered the basic optimal estimation problem, and our task for the remainder of this section consists in tidying up some loose ends, and presenting examples. Here, in order, is what we shall do.

1. We shall eliminate the restrictive Assumption 8.4-2, which required a zero plant input and zero mean initial state—the resulting change in the estimator is very minor. Then we shall present a simple example.
2. We shall present an interpretation of the matrix $P(t)$, the solution of the Riccati equation (8.4-15), as a measure of the goodness of the estimate at time t, and we shall present a second example exhibiting this interpretation.
3. We shall show how to cope with the case $t_0 = -\infty$, drawing special attention to time-invariant problems, and including one example.

Elimination of Assumption 8.4-2. We wish to consider situations where $u(t)$ can be nonzero, and $E[x(t_0)] = m$ can be nonzero. The effect of nonzero values of either of these quantities will be to leave the plant state covariance and output covariance the same as before, but to change the mean of the

plant state and plant output. Thus, since from (8.4-1),

$$x(t) = \Phi(t, t_0)x(t_0) + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)u(\tau)\, d\tau + \int_{t_0}^{t} \Phi(t, \tau)v(\tau)\, d\tau,$$

it follows that now

$$E[x(t)] = \Phi(t, t_0)m + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)u(\tau)\, d\tau. \tag{8.4-24}$$

A detailed analysis shows that $E[x_e(t)]$ will have the same value as $E[x(t)]$ if each entry of $x_e(t)$ is to be a minimum variance estimate of the corresponding entry of $x(t)$, but otherwise the covariance of $x_e(t)$ is unaltered. The way to achieve this is to make a modification to the estimator equations (8.4-21) and (8.4-23). (The required modification might even have been guessed from Chapter 8, Sec. 8.2.) Without carrying out a somewhat complicated derivation, we merely state the modified equations:

$$\frac{d}{dt}x_e(t) = F_e(t)x_e(t) - K_e(t)y(t) + G(t)u(t) \quad x_e(t_0) = m \tag{8.4-25}$$

or

$$\frac{d}{dt}x_e(t) = F(t)x(t) + K_e(t)[H'(t)x_e(t) - y(t)] + G(t)u(t) \quad x_e(t_0) = m, \tag{8.4-26}$$

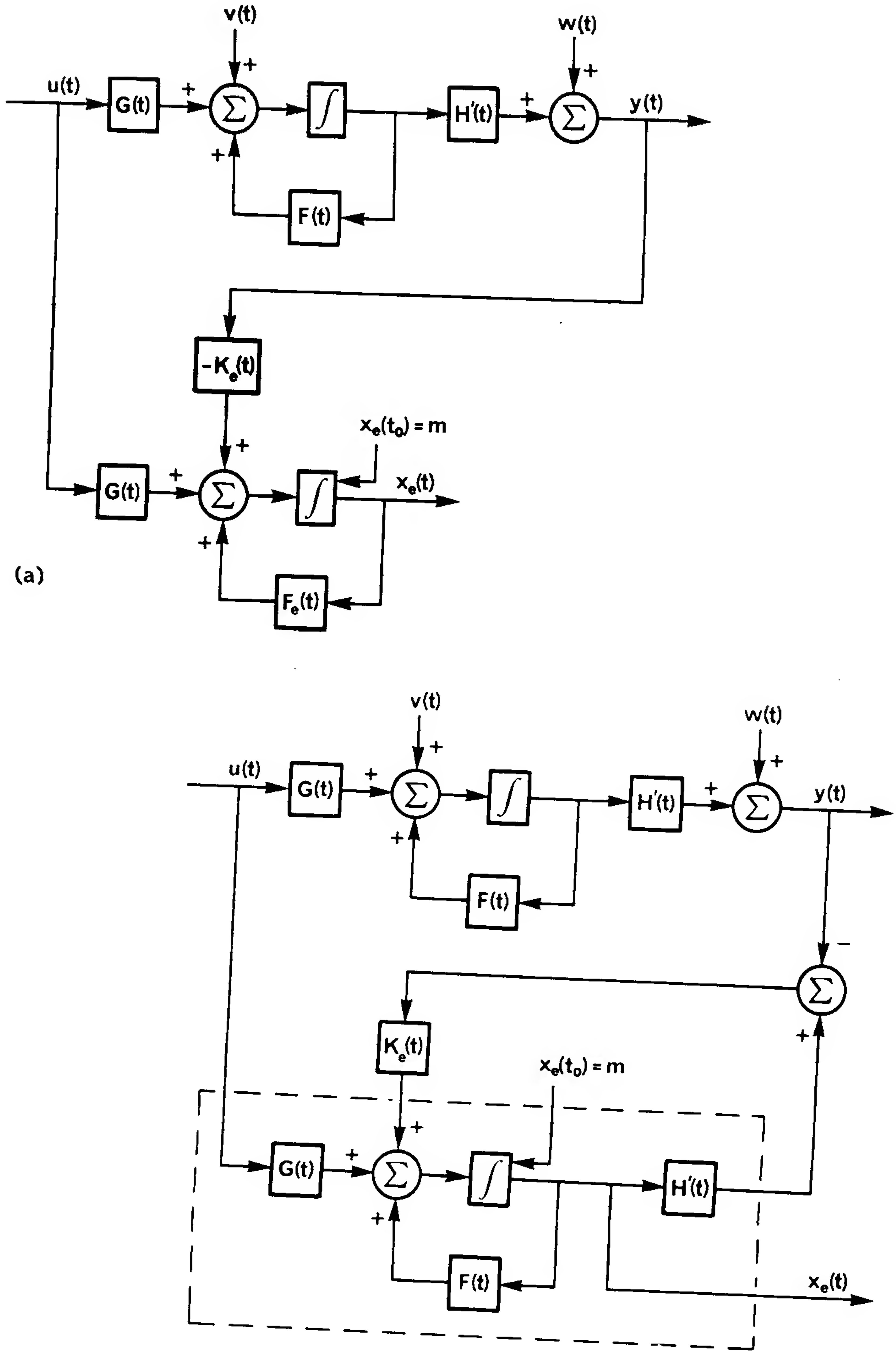
where, as before, $K_e(t)$ is given by (8.4-22) and $P(t)$ satisfies the Riccati equation (8.4-15). Figure 8.4-4(a) shows plant and estimator according to (8.4-25), and Fig. 8.4-4(b) shows plant and estimator according to (8.4-26). Part of the estimator in Fig. 8.4-4(b) is enclosed in dotted lines, to emphasize the fact that the estimator is a model of the plant with additions. Let us now summarize.

> ***Physical realization of the optimal estimator*** (***t_0 finite***). Given the plant equations (8.4-1) and (8.4-2), suppose that Assumptions 8.4-1 and 8.4-3 hold. Then an on-line estimate $x_e(t)$ of $x(t)$ is provided (at time t) by the arrangement of Fig. 8.4-4(a) [or 8.4-4(b)], where $K_e(t)$ is defined via (8.4-22) and $P(t)$ via (8.4-15). The matrix $P(t)$ exists for all $t \geq t_0$, and is symmetric nonnegative definite.

By way of example, we consider the following problem. The plant is time invariant, with transfer function $1/(s + 1)$, and there is input noise of covariance $1\delta(t - \tau)$, and additive noise at the output of covariance $2\delta(t - \tau)$. At time zero, the initial state has a known value of 1. The problem is to design an optimal state estimator. In state-space terms, the plant equations are

$$\dot{x} = -x + u + v \quad x(0) = 1$$

$$y = x + w$$



(b)

Fig. 8.4-4 (a) Full plant estimator structure; (b) redrawn estimator.

with $E[v(t)v(\tau)] = Q\delta(t-\tau) = 1\delta(t-\tau)$ and $E[w(t)w(\tau)] = R\delta(t-\tau) = 2\delta(t-\tau)$. The initial state covariance matrix P_0 is zero. From Eq. (8.4-15), we have

$$\dot{P} = -2P - \tfrac{1}{2}P^2 + 1.$$

This equation yields

$$\int_0^{P(t)} \frac{dP}{P^2 + 4P - 2} = -\frac{1}{2}\int_0^t dt$$

whence

$$\left[\frac{1}{2\sqrt{6}} \ln \frac{P + 2 - \sqrt{6}}{P + 2 + \sqrt{6}}\right]_0^{P(t)} = -\frac{1}{2}t$$

or

$$P(t) = \frac{(\sqrt{6} - 2)[1 - \exp(-\sqrt{6}\,t)]}{1 + [(\sqrt{6} - 2)/(\sqrt{6} + 2)]\exp(-\sqrt{6}\,t)}$$

The gain matrix $K_e(t)$ for the optimal estimator, here a scalar, is given from (8.4-22) as $\frac{1}{2}P(t)$. Figure 8.4-5 shows the plant and estimator.

The matrix $P(t)$ as a measure of goodness of estimation. It may well be asked whether $P(t)$ has any significance other than aiding in the computation of $K_e(t)$. Indeed it has, as we shall now show. Recall Eq. (8.4-10), rewritten for convenience as

$$E\left\{\left[b'x(t_1) - \int_{t_0}^{t_1} s'(t)y(t)\,dt\right]^2\right\}$$
$$= r'(t_0)P_0r(t_0) + \int_{t_0}^{t_1} [s'(t)R(t)s(t) + r'(t)Q(t)r(t)]\,dt. \qquad (8.4\text{-}10)$$

We recall that we found the function $s(\cdot)$ minimizing both sides of (8.4-10) by viewing the problem of minimizing the right side of (8.4-10), subject to (8.4-8), as an optimal regulator problem. We have not bothered to note till now what this minimum value is. A quick review of the discussion regarding the minimizing of the right side of (8.4-10) will show this minimum to be, in fact, $b'P(t_1)\,b$. Since $\int_{t_0}^{t_1} s'(t)y(t)\,dt$ is $b'x_e(t_1)$, it therefore follows that

$$E\{[b'x(t_1) - b'x_e(t_1)]^2\} = b'P(t_1)b.$$

But since b is arbitrary, it is evident that

$$E\{[x(t_1) - x_e(t_1)][x(t_1) - x_e(t_1)]'\} = P(t_1). \qquad (8.4\text{-}27)$$

Thus, $P(t_1)$ is a measure of the error between $x(t_1)$ and $x_e(t_1)$; more precisely, it is the error covariance.

The following example is discussed in [17]. The position and velocity of a satellite are to estimated; the measurements available are, first, a position measurement including white gaussian noise, and second, an acceleration measurement including a constant random variable measurement error, accounting for drift, etc. The motion of the satellite is linear and one-dimen-

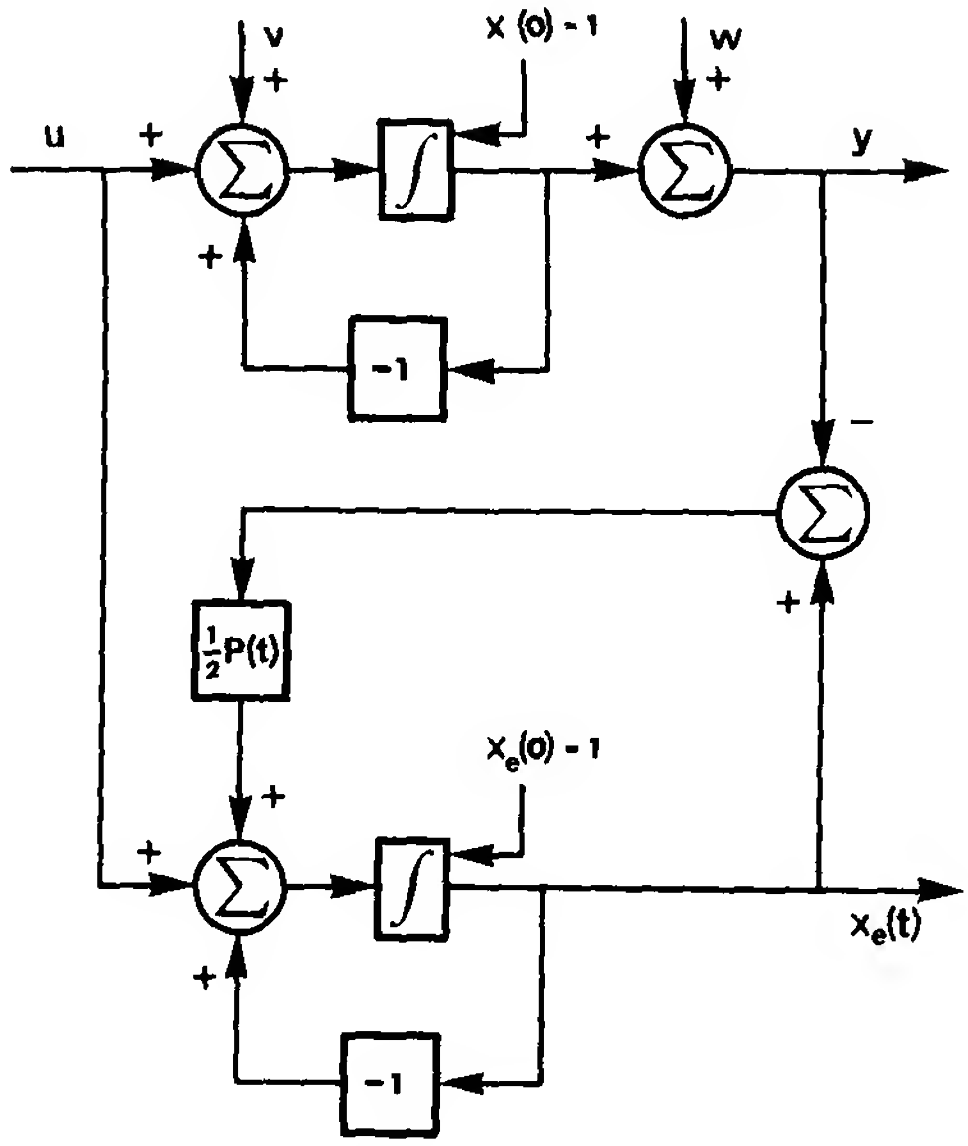


Fig. 8.4-5 Plant and optimal estimator.

sional, and there is a constant gaussian random acceleration, independent of the measurement error in the acceleration measurement.

With x_1 denoting position and x_2 velocity, the equations of motion are

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = a$$

where a is the constant acceleration and is a gaussian random variable. Now the problem data concerning measurements and noise do not immediately allow construction of the standard system equations; therefore, we proceed as follows. The measurement of acceleration is $y_2 = a + b$, where b is a constant gaussian random variable. Then

$$\dot{x}_2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} y_2 + \left(\frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2} y_2 - b\right)$$

where $\sigma_a = E[a^2]$ and $\sigma_b = E[b^2]$. Observe that $[\sigma_a^2/(\sigma_a^2 + \sigma_b^2)]y_2$ and $[\sigma_b^2/(\sigma_a^2 + \sigma_b^2)]y_2 - b$ are independent, for

$$\begin{aligned} E\left\{\left[\frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} y_2\right]\left[\frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2} y_2 - b\right]\right\} &= \frac{\sigma_a^2 \sigma_b^2}{(\sigma_a^2 + \sigma_b^2)^2} E(y_2^2) - \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} E(y_2 b) \\ &= \frac{\sigma_a^2 \sigma_b^2}{(\sigma_a^2 + \sigma_b^2)^2}(\sigma_a^2 + \sigma_b^2) - \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}\sigma_b^2 \\ &= 0. \end{aligned}$$

So this equation is of the form

$$\dot{x}_2 = u + x_3,$$

where u is known and independent of x_3. Moreover, x_3 is a constant gaussian random variable whose variance is easily checked to be $[\sigma_a^2\sigma_b^2/(\sigma_a^2 + \sigma_b^2)]$, a quantity that we shall call p from now on.

The full system equations thus become

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u$$

$$y = [1 \quad 0 \quad 0]x + w.$$

We assume that u is known and that the initial values of x_1 and x_2 are known, whereas $E[x_3^2(0)] = p$, and $E[w(t)w(\tau)] = r\delta(t - \tau)$.

The Riccati equation becomes

$$\dot{P} = P\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} P - P\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} r^{-1}[1 \quad 0 \quad 0]P$$

with initial condition

$$P(0) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & p \end{bmatrix}.$$

The solution to this equation turns out to be

$$P(t) = \frac{r}{t^5/20 + r/p}\begin{bmatrix} \dfrac{t^4}{4} & \dfrac{t^3}{2} & \dfrac{t^2}{2} \\ \dfrac{t^3}{2} & t^2 & t \\ \dfrac{t^2}{2} & t & 1 \end{bmatrix},$$

and the optimal gain vector is

$$k_e' = -\left[\frac{t^4}{4(t^5/20 + r/p)} \quad \frac{t^3}{2(t^5/20 + r/p)} \quad \frac{t^2}{2(t^5/20 + r/p)}\right].$$

In the limit as $t \to \infty$, $P(t) = 0$ and $k_e(t) = 0$. Essentially what happens is that x_3 is exactly identified as $t \to \infty$. Since $x_1(0)$ and $x_2(0)$ are both known, this means that x_1 and x_2 become exactly identified as $t \to \infty$. In consequence, the error covariance approaches zero—i.e., $P(t) \to 0$. Simultaneously, the noisy measurements become of no use, and thus $k_e(t) \to 0$.

Initial times in the infinite past. The interpretation of $P(t)$ as the error covariance will now be used in a discussion of the estimation problem for $t_0 = -\infty$. At this stage therefore, we drop Assumption 8.4-3, but introduce a new one.

ASSUMPTION 8.4-4. For all t, the pair $[F(t), H(t)]$ is completely observable.

To get a rough idea of the reason for this assumption, consider a time-invariant plant having all unobservable, unstable states. Suppose also that at time $t_0 = -\infty$, the initial state of the plant has mean zero. Since any estimator can deduce absolutely no information about the plant state from the available measurements, the only sensible estimate of the plant state is $x_e(t_1) = 0$. Now the covariance of the plant state vector will be infinite at any finite time, and therefore so will the error covariance. Consequently, if $P(t)$ retains the significance of being the error covariance in the $t_0 = -\infty$ case, it is infinite. The gain K_e of the optimal filter will certainly have some, if not all, entries infinite also. The purpose of Assumption 8.4-4 is to prevent this sort of difficulty. In fact, under this assumption, and, to simplify matters, under the assumption $P_0 = 0$, we claim that $P(t)$ exists as the solution of (8.4-15) (and is finite, of course) when (8.4-15) has the boundary condition $\lim_{t_0 \to -\infty} P(t_0) = 0$. A formal justification of this result can be achieved as follows. With $\hat{t} = -t$, $\hat{F}(\hat{t}) = F(t)$, and $\hat{H}(\hat{t}) = H(t)$, the complete observability of $[F(t), H(t)]$ implies the complete controllability of $[\hat{F}'(\hat{t}), \hat{H}(\hat{t})]$. Considering the regulator problem associated with (8.4-11) and (8.4-12), we see by our earlier results on the infinite-time regulator problem that the existence of a solution $\hat{P}$ to the Riccati differential equation (8.4-13) is guaranteed under the boundary condition $\lim_{\hat{t}_0 \to \infty} \hat{P}(\hat{t}_0) = 0$. Since $\hat{P}(\hat{t}) = P(t)$, the claim is established.

If F and H are constant (i.e., the plant is time invariant) and if Q and R are constant (i.e., the noise is stationary), it follows (again, by examining the dual regulator problem) that the value of $P(t)$ obtained by letting $t_0 \to -\infty$ is independent of time, and can therefore be computed by evaluating $\lim_{t \to \infty} P(t)$ where $P(t)$ satisfies (8.4-15) with the initial condition $P(0) = 0$. Also, the constant matrix P is a solution of the quadratic matrix equation

$$PF' + FP - PHR^{-1}H'P + Q = 0. \tag{8.4-28}$$

The gain of the optimal estimator is then constant, being given by

$$K_e = -PHR^{-1} \tag{8.4-29}$$

and if G is constant, the optimal estimator is a time-invariant system:

$$\dot{x}_e = Fx_e + Gu + K_e[H'x_e - y]. \tag{8.4-30}$$

However, this equation is of little importance from the practical point of view unless it represents an asymptotically stable system or, equivalently, $F_e + K_eH'$ has eigenvalues with negative real parts. The way to ensure this is to require the following.

ASSUMPTION 8.4-5. With constant F and Q, and D any matrix such that $DD' = Q$, the pair $[F, D]$ is completely controllable.

One way to prove the asymptotic stability of the optimal estimator under this assumption (with also G, H, and R constant, $t_0 = -\infty$), is to

examine the associated regulator problem and to apply the results of earlier chapters. The assumption serves to provide an observability assumption for the regulator problem, which guarantees asymptotic stability for the optimal regulator. This carries over to the optimal estimator.

For the time-invariant plant, stationary noise problem, if Assumptions 8.4-4 and 8.4-5 hold, it can also be shown that P is positive definite and is the unique positive definite solution of (8.4-28). Since the time-invariant problem is so important, we shall summarize the results.

> ***The time-invariant optimal estimator.*** Given the plant equations (8.4-1) and (8.4-2), suppose that F, G, and H are constant. Suppose also that Assumptions 8.4-1 and 8.4-4 hold, that Q and R are constant, and that $t_0 = -\infty$. Then the matrix P, which is the error covariance, is constant and satisfies (8.4-28); it may be computed by taking a limiting boundary condition of the form $\lim_{t_0 \to -\infty} P(t_0) = 0$ for (8.4-15) and evaluating $P(t)$ for any t, or by evaluating $\lim_{t \to \infty} P(t)$ where $P(t)$ satisfies (8.4-15) but with boundary condition $P(0) = 0$. The optimal estimator is time invariant. Moreover, if Assumption 8.4-5 holds, the optimal estimator is asymptotically stable.

The result is, of course, closest to the ideas of Sec. 8.2. There, we confined our attention to time-invariant systems, and time-invariant estimators. There, also, the estimators were asymptotically stable, as a result of the complete observability assumption.

Let us now consider an example. We suppose white noise of covariance $q\delta(t - \tau)$ is the input to a plant of transfer function $1/s(s + 1)$, starting from time $t_0 = -\infty$. There is additive output noise of covariance $r\delta(t - \tau)$. Thus, we take

$$F = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \quad g = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad h' = [0 \quad 1].$$

Figure 8.4-6(a) shows the scheme. To put this in the standard form, we set

$$E[v(t)v'(\tau)] = \begin{bmatrix} q & 0 \\ 0 & 0 \end{bmatrix}.$$

The input noise is not available to the estimator as an input. Denoting by p_{ij} the entries of the P matrix, the quadratic equation (8.4-28) becomes

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}$$
$$- \frac{1}{r}\begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}[0 \quad 1]\begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix} + \begin{bmatrix} q & 0 \\ 0 & 0 \end{bmatrix} = 0$$

or

$$-\frac{1}{r}p_{12}^2 + q = 0$$

$$p_{11} - p_{12} - \frac{1}{r}p_{12}p_{22} = 0$$

$$2p_{12} - 2p_{22} - \frac{1}{r}p_{22}^2 = 0.$$

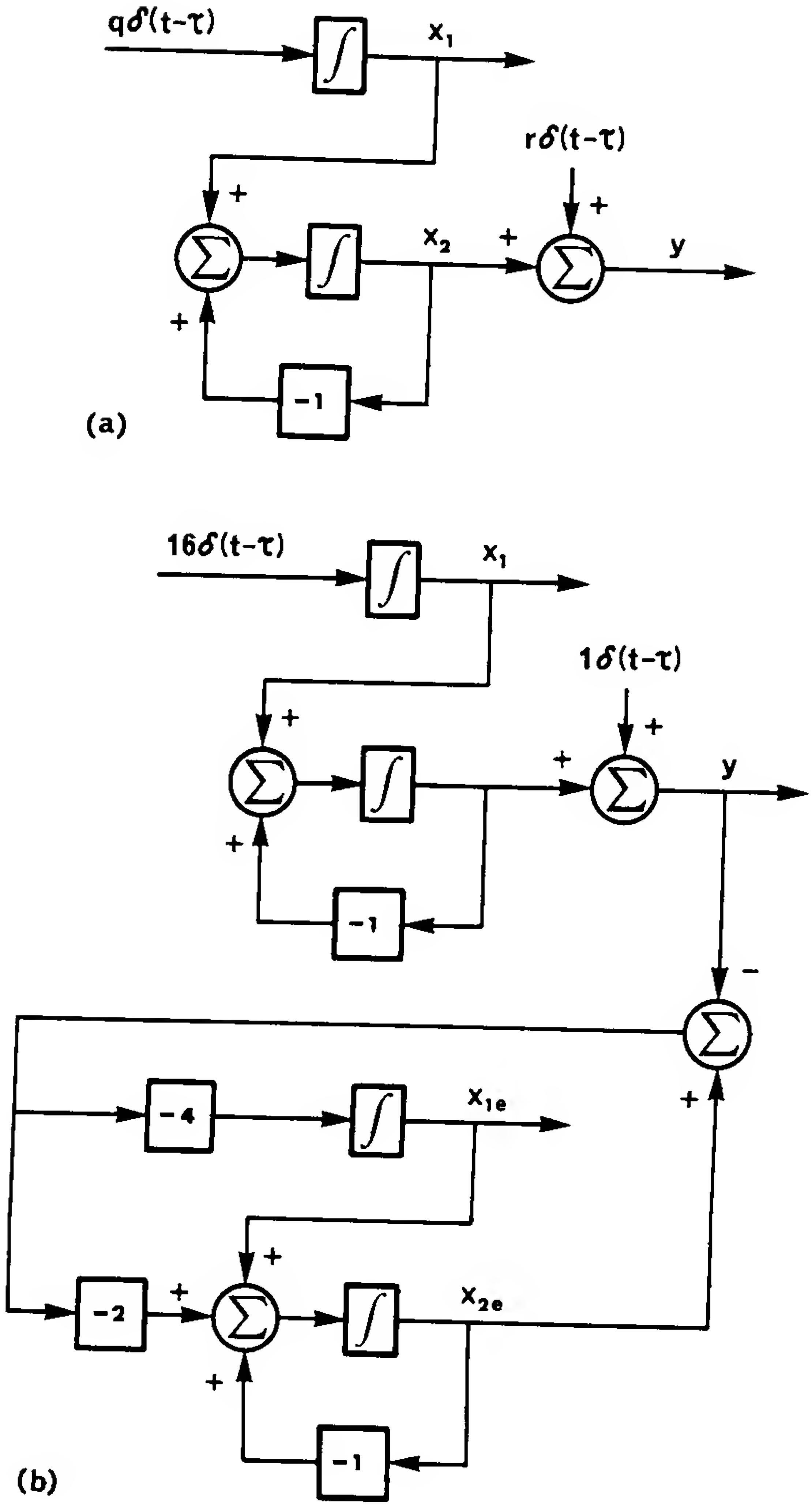


Fig. 8.4-6 A specific plant and state estimator.

It follows that $p_{12} = \sqrt{rq}$ from the first equation, $p_{22} = r[\sqrt{1 + 2\sqrt{q/r}} - 1]$ from the third equation, and $p_{11} = \sqrt{rq + 2q}$ from the second equation. In solving the third equation, only the positive solution for p_{22} is acceptable, since P must be positive definite. The estimator gain from Eq. (8.4-29) is

$$k_e = \begin{bmatrix} -\sqrt{q/r} \\ -\sqrt{1 + 2\sqrt{q/r}} + 1 \end{bmatrix}.$$

If, e.g., $q = 16$, $r = 1$, we have $k_e' = [-4 \quad -2]$. The matrix F_e is $F + k_e h'$, or

$$\begin{bmatrix} 0 & -4 \\ 1 & -3 \end{bmatrix}.$$

It is readily checked to have eigenvalues with negative real parts. The plant and estimator are shown in Fig. 8.4-6(b) for the case $q = 16$, $r = 1$.

We conclude this section with two comments. First, an optimally designed estimator may be optimal for noise covariances differing from those assumed in its design. This holds for precisely the same reason that a control law resulting from an optimal design may be optimal for more than one performance index, a point we discussed earlier.

Second, in the interests of achieving an economical realization of an estimator, it may be better to design a suboptimal one that is time invariant, rather than an optimal one that is time varying. For example, suppose the plant whose states are being estimated is time invariant, and that estimation is to start at time zero. Suppose also that at this time, the initial state of the plant is known to be zero. Then there would not normally be a great deal of loss of optimality if, instead of implementing the optimal estimator which would be time varying, a time-invariant estimator were implemented, designed perhaps on the assumption that the initial time t_0 were $-\infty$ and not zero. Reference [8] obtains quantitative bounds on the loss of optimality incurred in this arrangement.

Problem 8.4-1. Consider the plant $\dot{x} = x + v$, $y = x + w$, with $E[v(t)v'(\tau)] = E[w(t)w'(\tau)] = \delta(t - \tau)$ and v and w independent. Suppose that at time zero, $x(0)$ is known to be zero. Design an optimal estimator.

Problem 8.4-2. Repeat Problem 8.4-1 for the case when the initial time t_0 is $-\infty$. Also with $t_0 = -\infty$, suppose that an additional measurement becomes available, i.e., suppose now

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

where $y_1 = x + w_1$, $y_2 = x + w_2$ and $E[w_1(t)w_1(\tau)] = E[w_2(t)w_2(\tau)] = \delta(t - \tau)$

with v, w_1 and w_2 independent. Design an optimal estimator for this case. Compare the error covariances for the single and multiple output cases.

Problem 8.4-3. Suppose that $\dot{x} = ax + v$, $y_1 = x + w_1$, $y_2 = x + w_2$, where v, w_1, and w_2 all are independent, with covariances $q\delta(t-\tau)$, $r_{11}\delta(t-\tau)$, and $r_{22}\delta(t-\tau)$, respectively. For the $t_0 = -\infty$ case, derive analytic expressions for the error covariance, assuming that y_1 alone is available and that both y_1 and y_2 are available.

Problem 8.4-4. Given $\dot{x} = Fx + Gu + v$, $y = H'x + w$, with F, G, and H constant and v and w stationary white noise, suppose that t_0 is finite. Show that the estimator will be time invariant if $E[x(t_0)x'(t_0)]$ takes on a particular value—in general, nonzero.

Problem 8.4-5. Consider the standard estimation problem with $u(t) \equiv 0$, $E[x(t_0)] = 0$, and t_0 finite, save that $v(\cdot)$ and $w(\cdot)$ are no longer independent: rather, $E[v(t)w'(\tau)] = S(t)\delta(t-\tau)$ for some matrix $S(t)$. Show that the problem of finding a minimum variance estimate of $b'x(t_1)$ for arbitrary constant b and arbitrary t_1 is again equivalent to a quadratic regulator problem, with a cross-product term between state and control in the loss function. Attempt to solve the complete estimation problem.

Problem 8.4-6. (The smoothing problem.) Let $x_e(t_0 \mid t_1)$ denote the minimum variance estimate of $x(t_0)$, given measurements $y(t)$ up till time $t_1 \geq t_0$, where, as usual, $\dot{x} = Fx + Gu + v$, $y = H'x + w$, $E[v(t)v'(\tau)] = Q(t)\delta(t-\tau)$, $E[w(t)w'(\tau)] = R(t)\delta(t-\tau)$, $E[x(t_0)x'(t_0)] = P_0$, $E[x_0] = m$, and v, w, and $x(t_0)$ are independent and gaussian, the first two also being zero mean. It is desired to define a procedure for computing $x_e(t_0 \mid t_1)$.

Consider the scheme of Fig. 8.4-7. Observe that $x_{1e}(t \mid t) = x_e(t \mid t)$ and $x_{2e}(t \mid t) = x_e(t_0 \mid t)$, implying that the smoothing problem may be viewed as a filtering problem. Show that

$$\dot{x}_{1e} = (F - P_{11}HR^{-1}H')x_{1e} + P_{11}HR^{-1}y \qquad x_{1e}(t_0 \mid t_0) = m$$

$$\dot{x}_{2e} = -P'_{12}HR^{-1}H'x_{1e} + P'_{12}HR^{-1}y \qquad x_{2e}(t_0 \mid t_0) = m$$

where

$$\dot{P}_{11} = P_{11}F' + FP_{11} - P_{11}HR^{-1}H'P_{11} + Q \qquad P_{11}(t_0) = P_0$$

$$\dot{P}_{12} = (F - PHR^{-1}H')P_{12}. \qquad P_{12}(t_0) = P_0$$

Show also how to find $x_e(t_2 \mid t_1)$ for arbitrary $t_2 < t_1$. [*Hint:* Represent the system of Fig. 8.4-7 as a $2n$-dimensional system with F matrix F_a, etc., and let

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P'_{12} & P_{22} \end{bmatrix}$$

be the solution of the associated filtering Riccati equation. Use P to define the optimal $2n$-dimensional filter, and show that the two differential equations shown follow from the optimal filter description. This technique appeared in [15].]

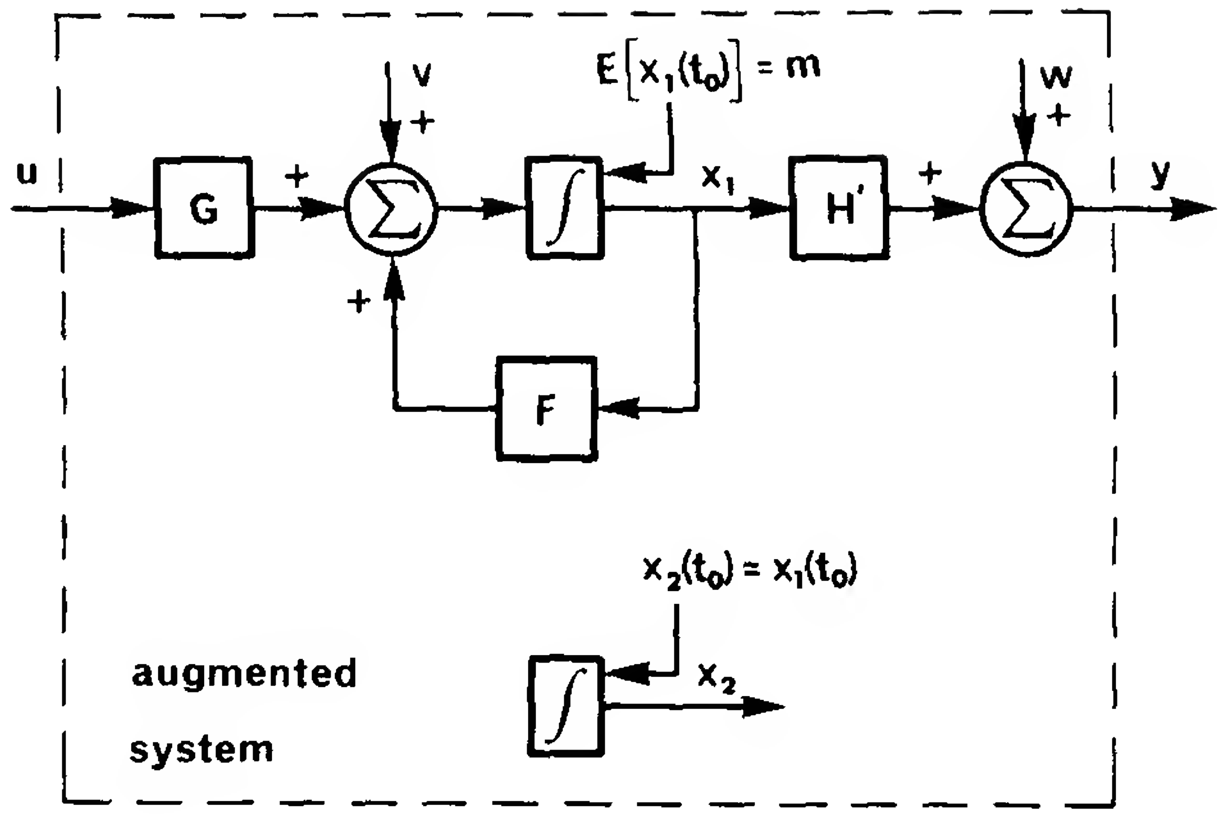


Fig. 8.4-7 System augmented with integrators.

REFERENCES

[1] Popov, V. M., "Hyperstability and Optimality of Automatic Systems with Several Control Functions," *Rev. Roumaine Sci. Tech.—Electrotechnol. Energ.*, Vol. 9, No. 4, 1964, pp. 629–690.

[2] Anderson, B. D. O., and D. G. Luenberger, "Design of Multivariable Feedback Systems," *Proc. IEE*, Vol. 114, No. 3, March 1967, pp. 395–399.

[3] Luenberger, D. G., "Canonical Forms for Linear Multivariable Systems," *IEEE Trans. Auto. Control*, Vol. AC–12, No. 3, June 1967, pp. 290–293.

[4] Wonham, W. M., "On Pole Assignment in Multi-Input Controllable Linear Systems," *IEEE Trans. Auto. Control*, Vol. AC–12, No. 6, December 1967, pp. 660–665.

[5] Tuel, W. G., "On the Transformation to (Phase-Variable) Canonical Form," *IEEE Trans. Auto. Control*, Vol. AC–11, No. 3, July 1966, p. 607.

[6] Gopinath, B., and B. O. Lange, "On the Identification and Control of Linear Systems," *Tech. Rep. SUDAAR No. 351*, June 1968, Dept. of Aeronautics and Astronautics, Stanford Univ., Calif.

[7] Rosenbrock, H. H., "The Assignment of Closed Loop Poles," *Control Systems Centre Rep. No. 20*, October 1967, Control Systems Centre, University of Manchester, England.

[8] Singer, R. A., "The Design and Synthesis of Linear Multivariable Systems with Application to State Estimation," *Tech. Rep. 6302–8*, June 1968, Stanford Electronics Laboratories, Stanford, Calif.

[9] Luenberger, D. G., "Observers for Multivariable Systems," *IEEE Trans. Auto. Control*, Vol. AC–11, No. 2, April 1966, pp. 190–197.

[10] Luenberger, D. G., "Observing the State of a Linear System," *IEEE Trans. Military Electron.*, Vol. MIL–8, No. 2, April 1964, pp. 74–80.

[11] Wiener, N., "*Extrapolation, Interpolation and Smoothing of Stationary Time Series*," M. I. T. Press, Cambridge, Mass., 1949.

[12] Bucy, R. S., and P. D. Joseph, *Filtering for Stochastic Processes and Applications to Guidance*, Interscience Publishers, New York, 1968.

[13] Kalman, R. E., and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 83, March 1961, pp. 95–108.

[14] Bryson, A. E., and M. Frazier, "Smoothing for Linear and Nonlinear Dynamic Systems," *Proc. Optimum Systems Synthesis Conf.*, USAF Tech. Rep. ASD–TDR–063–119, February 1963.

[15] Zachrisson, L. E., "On Optimal Smoothing of Continuous Time Kalman Processes," *Inform. Sci.*, Vol. 1, 1969, pp. 143–172.

[16] Bryson, A. E., and D. E. Johansen, "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise," *IEEE Trans. Auto. Control*, Vol. AC–10, No. 1 January 1965, pp. 4–10.

[17] Kalman, R. E., and T. S. Englar, "A User's Manual for the Automatic Synthesis Program," *NASA Contractor Rep. NASA CR–475*, June 1966.

CHAPTER 9

SYSTEM DESIGN USING STATE ESTIMATORS

9.1 CONTROLLER DESIGN—BASIC VERSIONS

This chapter is concerned with tying together the notions of state-variable feedback and estimation. In other words, we consider controllers of the sort shown in Fig. 9.1-1. In this section, we describe the basic controllers obtained by using the estimator designs of the previous chapter, and we discuss various properties of the overall controlled system. These properties include frequency domain characteristics, stability properties, and the carrying over of the various properties that optimality implies. The remaining sections complement and extend the ideas, principally by pointing out modified versions of the controller.

As our basic plant, we take the time-invariant system

$$\dot{x} = Fx + Gu \tag{9.1-1}$$

$$y = H'x. \tag{9.1-2}$$

For the observer, we shall take for the moment the structure of Chapter 8, Secs. 8.2 and 8.4—viz.,

$$\dot{x}_e = (F + K_e H')x_e + Gu - K_e y. \tag{9.1-3}$$

We assume that K_e is chosen so that all eigenvalues of $F + K_e H'$ have negative real parts. Whether it is optimal for some noise statistics is irrelevant for our considerations here; either the scheme of Sec. 8.2 or that of Sec. 8.4 can

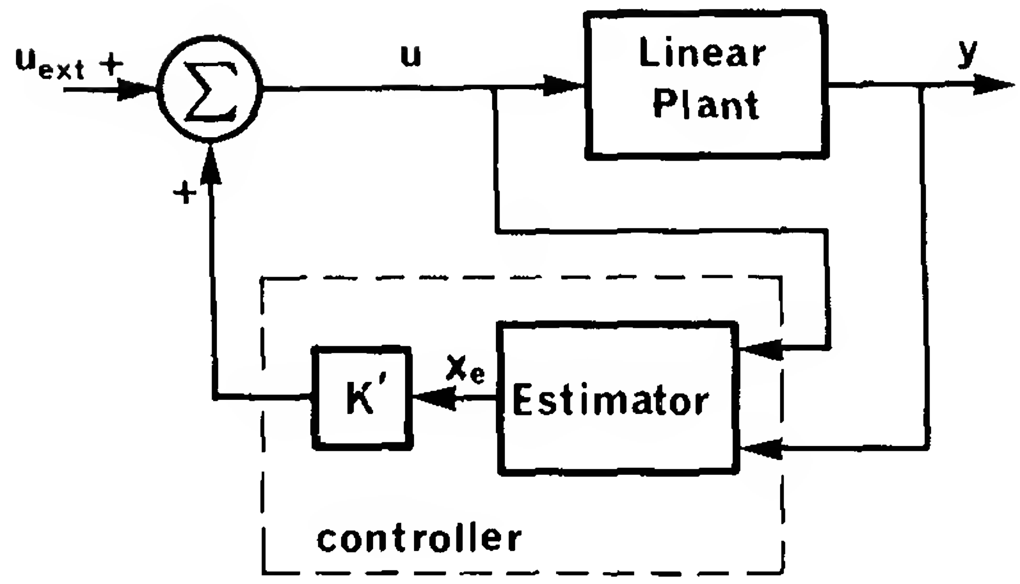


Fig. 9.1-1 The basic controller.

be assumed to lead to the choice of K_e. Subsequently, in this section we shall consider the use of Luenberger observers in controller design.

We further assume that we should like to implement the control law $u = u_{ext} + K'x$ (u_{ext} denoting an external input), where, presumably, K has been selected as an optimal control law. However, for obvious reasons, we implement the following law instead:

$$u = u_{ext} + K'x_e. \tag{9.1-4}$$

Equations (9.1-1) through (9.1-4) sum up the entire plant-controller arrangement. We shall now study some properties of the arrangement using these equations. We begin by making a slight rearrangement. First, from (9.1-1) and (9.1-4), we have

$$\dot{x} = (F + GK')x - GK'(x - x_e) + Gu_{ext}. \tag{9.1-5}$$

Second, from (9.1-1), (9.1-2), and (9.1-3), we have

$$\frac{d}{dt}(x - x_e) = (F + K_eH')(x - x_e), \tag{9.1-6}$$

which holds independently of u_{ext}. Now we regard the $2n$ vector, whose first n entries are x and whose second n entries are $x - x_e$, as a new state vector for the overall plant-controller scheme. (It would, of course, be equally valid to take as a state vector a $2n$ vector with the first n entries consisting of x and the second n entries consisting of x_e.)

The plant-controller arrangement, then, has the following description—the first equation following from (9.1-5) and (9.1-6), the second from (9.1-2):

$$\frac{d}{dt}\begin{bmatrix} x \\ x - x_e \end{bmatrix} = \begin{bmatrix} F + GK' & -GK' \\ 0 & F + K_eH' \end{bmatrix}\begin{bmatrix} x \\ x - x_e \end{bmatrix} + \begin{bmatrix} G \\ 0 \end{bmatrix} u_{ext} \tag{9.1-7}$$

$$y = [H' \quad 0]\begin{bmatrix} x \\ x - x_e \end{bmatrix}. \tag{9.1-8}$$

With input u_{ext} and output y, the plant-controller arrangement has the fol-

lowing transfer function matrix, derivable by manipulating (9.1-7) and (9.1-8):

$$W(s) = H'[sI - (F + GK')]^{-1}G. \tag{9.1-9}$$

This is exactly the transfer function matrix that would have resulted if true state-variable feedback were employed. The poles of the open-loop plant, corresponding to the zeros of det $(sI - F)$, are shifted to the zeros of det $[sI - (F + GK')]$. The zeros of a scalar $W(s)$ are unaltered.

Thus, from the steady-state (or zero initial state) point of view, use of the estimator as opposed to use of true state-variable feedback makes no difference. This is, of course, what should be expected. For the case in which the steady state has been reached, $x - x_e$ has approached zero and $x = x_e$, or, in the case of zero initial state, $x = 0$ and $x - x_e = 0$, so that again $x = x_e$. Clearly, with $x = x_e$, the control used is precisely that obtained with true state-variable feedback.

From the transient point of view, the plant-controller scheme of Eqs. (9.1-7) and (9.1-8) will behave differently from a scheme based on true state-variable feedback. Equation (9.1-7) defines a $2n$-dimensional system, whereas state-variable feedback yields an n-dimensional system. However, the $2n$-dimensional system is still asymptotically stable: inspection of (9.1-7) shows that the eigenvalues of $F + GK'$ and of $F + K_eH'$ determine the characteristic modes. (The eigenvalues of $F + K_eH'$ are, of course, associated with the additional new modes, which are evidently uncontrollable.)

We shall now consider a simple example, stemming from a second-order position controller. A servomotor, often through gears, drives an inertial load with coulomb friction. The transfer function relating the input voltage of the motor to the output shaft position is of the form $K/s(s + a)$, and here we shall assume normalization to $1/s(s + 1)$. This transfer function is representable by the state-space equations

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \quad y = [1 \quad 0]x.$$

We are required to choose a control law that minimizes $\int_{t_0}^{\infty} (u^2 + x_1^2 + x_2^2)\, dt$, and implement it with a state estimator, the associated state estimator poles being both at $s = -3$. First, the control law is obtained. In the usual notation,

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and $R = [1]$. We seek $\bar{P}$ as the positive definite solution of

$$\bar{P}F + F'\bar{P} - \bar{P}gR^{-1}g'\bar{P} + Q = 0.$$

By setting

$$\bar{P} = \begin{bmatrix} \bar{p}_{11} & \bar{p}_{12} \\ \bar{p}_{12} & \bar{p}_{22} \end{bmatrix}$$

and by substituting for F, g, Q, and R, we obtain the equations

$$\bar{p}_{12}^2 = 1$$
$$\bar{p}_{11} - \bar{p}_{12} - \bar{p}_{12}\bar{p}_{22} = 0$$
$$2\bar{p}_{12} - 2\bar{p}_{22} - \bar{p}_{22}^2 + 1 = 0.$$

The first equation gives $\bar{p}_{12}$, the third $\bar{p}_{22}$, and the second $\bar{p}_{11}$. Although there are multiple solutions, only one yields the following positive definite $\bar{P}$:

$$\bar{P} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

The optimal control law is now

$$u = u_{\text{ext}} + k'x$$

where $k' = -\bar{P}g$. Thus,

$$u = u_{\text{ext}} - x_1 - x_2.$$

We must now design an estimator, which will have an equation of the form

$$\dot{x}_e = (F + k_e h')x_e + gu - k_e y.$$

The problem is to choose a gain vector k_e so that $F + k_e h'$ has two eigenvalues at -3. Using the techniques of the previous chapter, or even by trial and error, we derive $k_e' = [-2 \quad -7]$. This leads to the plant-controller arrangement of Fig. 9.1-2.

We consider now how the ancillary properties of optimality may be derived. Rather than proceed exhaustively through an analysis of all of them, we shall indicate by considering only some properties the sort of thing that can be expected.

Figure 9.1-3(a) shows a nominally optimal system (with state estimation included in the control arrangement), where a nonlinearity $\phi(\cdot)$ has been introduced. For the zero external input case (i.e., for consideration of Lyapunov stability), the scheme of Fig. 9.1-3(b) is equivalent. The nonlinearity $\phi(\cdot)$ is assumed to be of the usual sector type—i.e., for a single-input system, the graph of the nonlinearity lies strictly within a sector bounded by straight lines of slope $\frac{1}{2}$ and ∞ (see Fig. 9.1-4). The appropriate generalization applies for multiple-input systems. Denoting the gain of the nonlinearity by β, we have $\frac{1}{2} < \beta < \infty$. Also, Eq. (9.1-7) is replaced by (for a single-input plant)

$$\frac{d}{dt}\begin{bmatrix} x \\ x - x_e \end{bmatrix} = \begin{bmatrix} F + \beta g k' & -\beta g k' \\ 0 & F + K_e H' \end{bmatrix}\begin{bmatrix} x \\ x - x_e \end{bmatrix} \tag{9.1-10}$$

where the argument of β—viz., $k'x$—has been suppressed. As before, $x - x_e$ decays exponentially. The equation for x yields

$$\dot{x} = (F + \beta g k')x + \beta w \tag{9.1-11}$$

where w is a vector all of whose entries decay exponentially to zero. By insist-

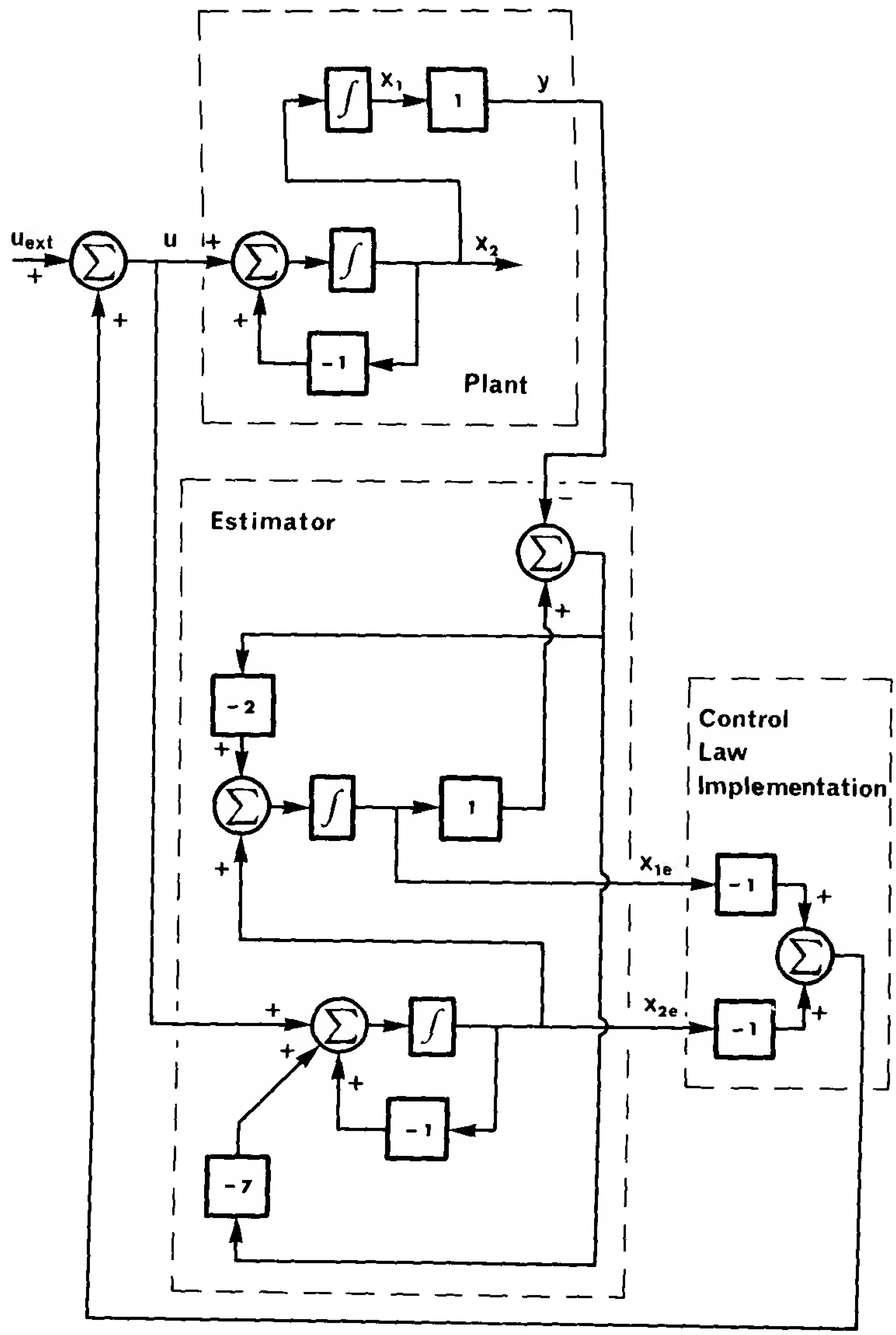


Fig. 9.1-2 Example of controller design.

ing that $\beta < \bar{\beta} < \infty$ for some constant $\bar{\beta}$, we can guarantee an exponential bound on βw. It turns out, although we shall not prove it, that the asymptotic stability properties of

$$\dot{x} = (F + \beta g k')x \tag{9.1-12}$$

established earlier then serve to guarantee also the asymptotic stability of (9.1-11).

It may well be argued that a more natural place to locate the nonlinearity

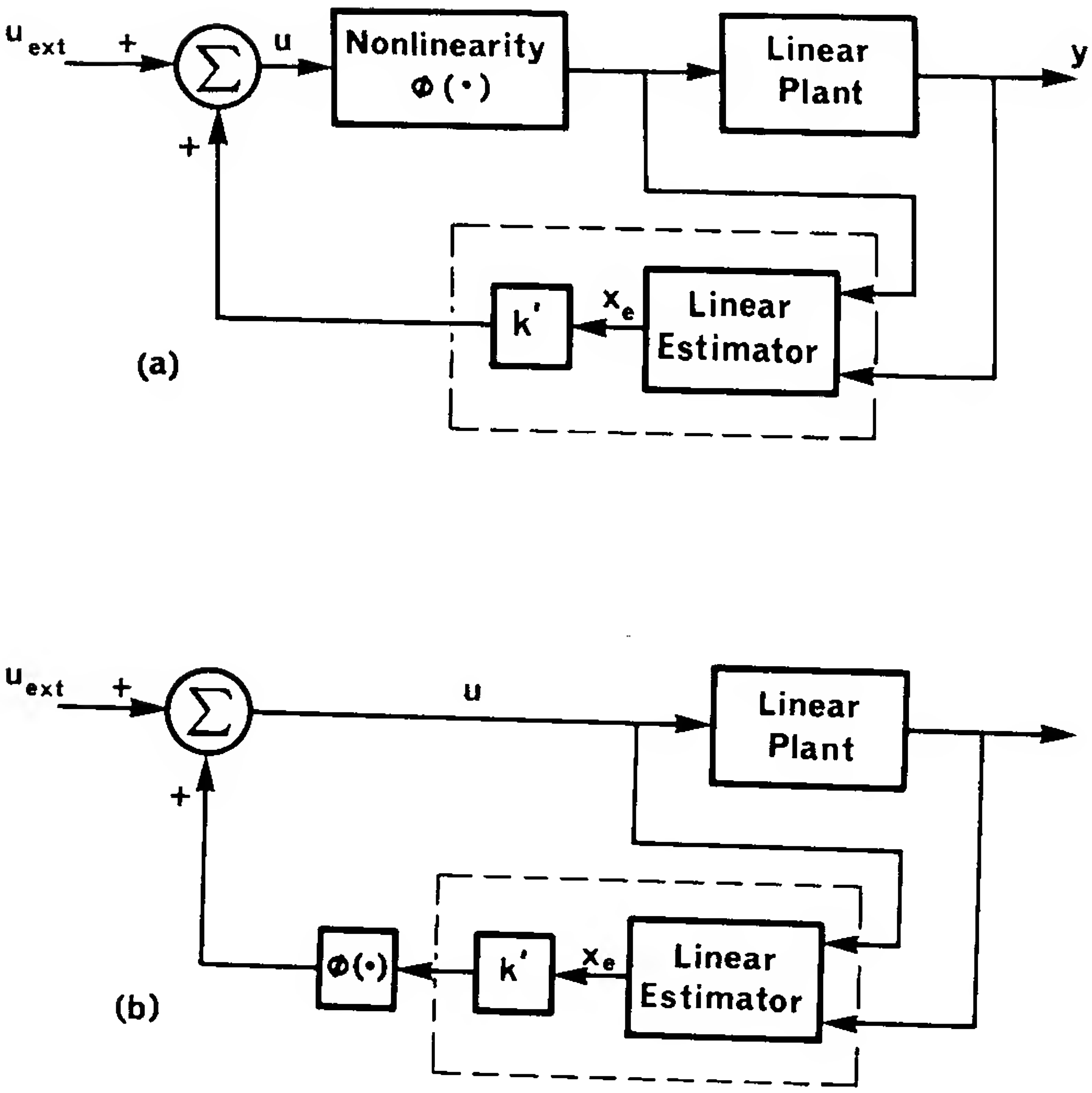


Fig. 9.1-3 Introduction of nonlinearity.

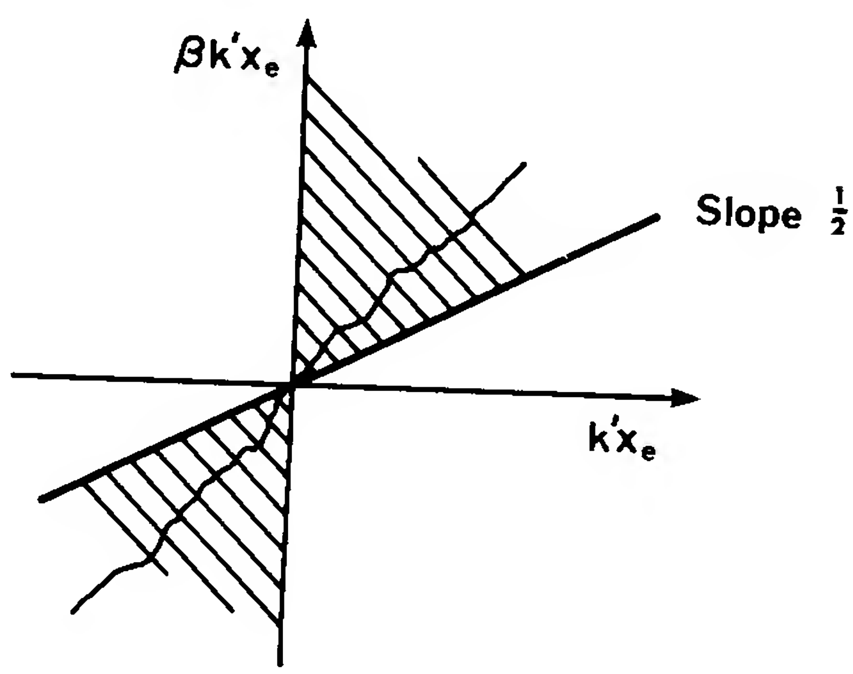


Fig. 9.1-4 A permissible nonlinearity.

would be right at the plant input, so that Fig. 9.1-5(a) would apply. If the estimator equation is still Eq. (9.1-3), repeated for convenience,

$$\dot{x}_e = (F + K_e H')x_e + gu - K_e y \tag{9.1-3}$$

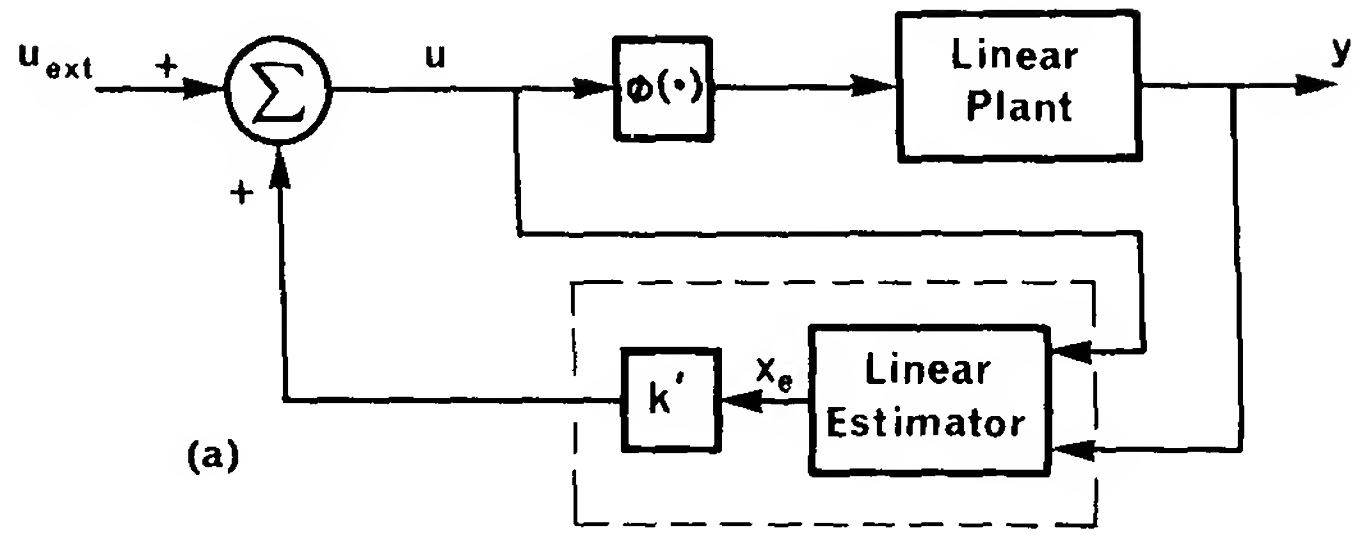
then, with plant equation

$$\dot{x} = Fx + \beta gu \tag{9.1-13}$$

we obtain

$$\frac{d}{dt}(x - x_e) = (F + K_e H')(x - x_e) + (\beta - 1)gu. \tag{9.1-14}$$

In this instance, x_e will not longer be a true estimate of x, unless u is zero.



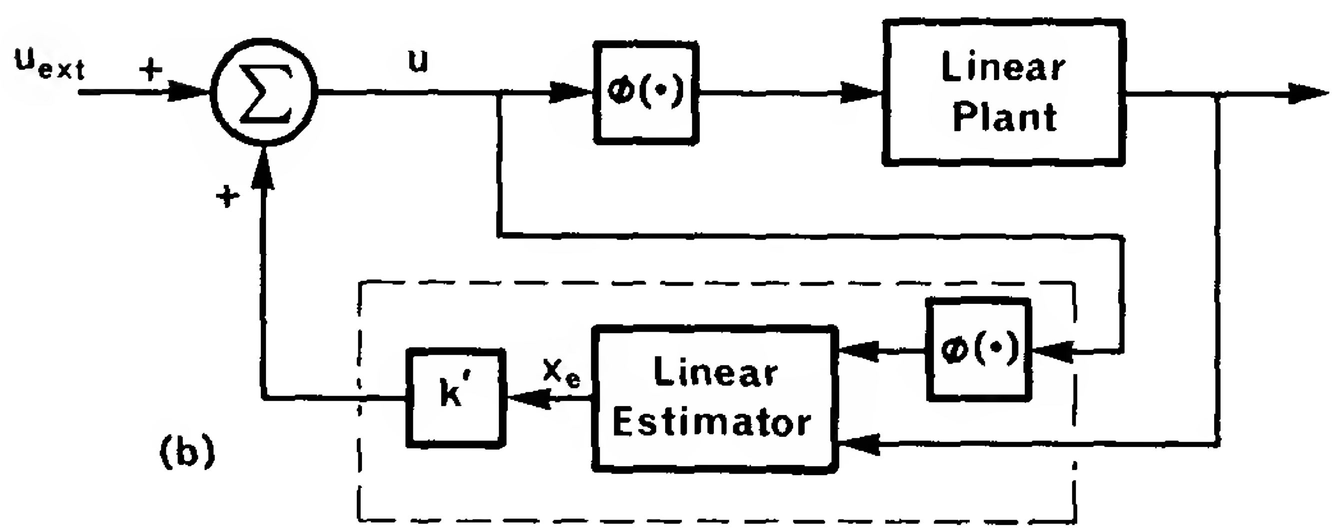


Fig. 9.1-5

Even if u decays exponentially, the rate at which x approaches x_e is no longer governed solely by the eigenvalues of $(F + K_eH')$. Something is seriously wrong, and, in general, the overall plant controller could not be expected to be asymptotically stable for all sector-limited β.

The only way out of this dilemma is to introduce the same nonlinearity into the controller, so that the estimator part of the controller again becomes a true model of the plant. [See Fig. 9.1-5(b).] The estimator equation (i.e., equation of the block marked "linear estimator" *together with* the nonlinearity) is

$$\dot{x}_e = (F + K_eH')x_e + \beta gu - K_ey. \tag{9.1-15}$$

Now, (9.1-13) and (9.1-15) together yield again

$$\frac{d}{dt}(x - x_e) = (F + K_eH')(x - x_e). \tag{9.1-3}$$

Therefore, correct estimation now takes place. Consequently, nonlinearities of the usual sector type will again be tolerable, in the sense that asymptotic stability will be retained.

However, there is a potential difficulty with this somewhat artificial arrangement. If the exact nature of the input nonlinearity to the plant is unknown, then it is impossible to construct the estimator. Of course, if the nature of the nonlinearity is roughly known, and the nonlinearity included

in the estimator approximates that in the plant, then, presumably, performance will be satisfactory.

As a second and representative property of optimal systems, we choose sensitivity reduction to parameter variations. We examine here the way this property carries over to the situation where an estimated, rather than an actual, state vector is used in implementing the optimal linear feedback law.

Figure 9.1-6 shows a redrawing of the plant-controller arrangement of Fig. 9.1-1, as a unity positive feedback system. The quantity fed back is $K'x_e$, which in the steady state becomes $K'x$. Therefore, the transfer function matrix of the block enclosed in dotted lines must be $K'(sI - F)^{-1}G$. This is precisely the same transfer function that arises when true state feedback is employed. We conclude that the same sensitivity results hold for the scheme of Fig. 9.1-6 as for the scheme when true state feedback is applied. More precisely, the output of the scheme of Fig. 9.1-6 will vary less as the result of a parameter change in the block in dotted lines than it would if a cascade controller were used to achieve the same transfer function. Moreover, with appropriate restrictions on the parameter variations, all components of the state of the linear plant part of the block, and the output of the linear plant, will have lower sensitivity to parameter change than the equivalent open-loop system.

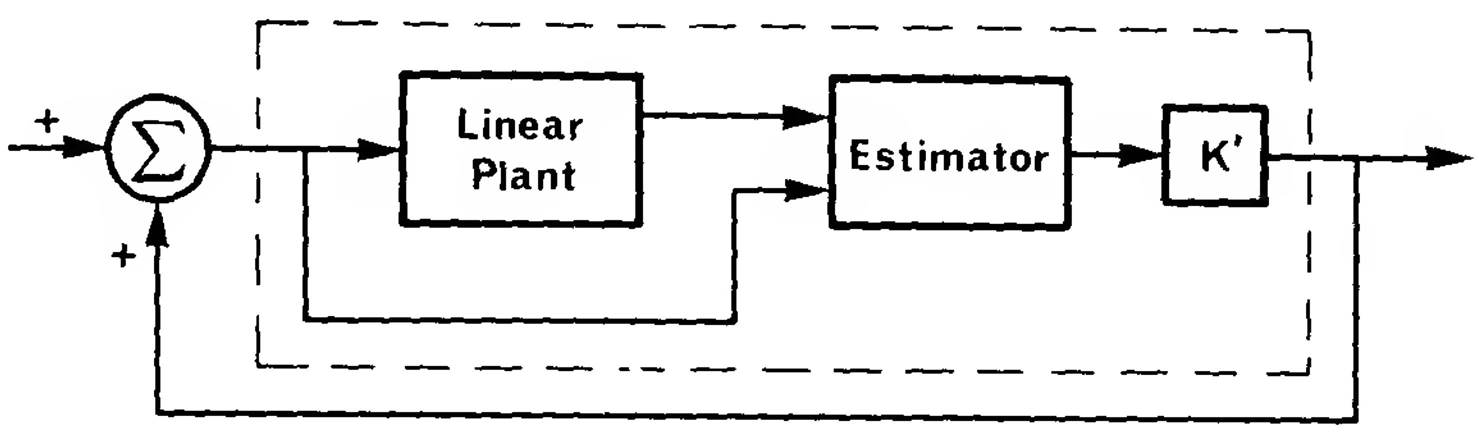


Fig. 9.1-6 Redrawing of the plant-controller arrangement.

In Chapter 7, we drew attention to the fact that the use of a feedback law achieving sensitivity reduction was tantamount to using an optimal feedback law. It might therefore be thought that use of the feedback law $u = K'x_e$ for the system defined by (9.1-1) through (9.1-3) might not then be merely *approximately* optimal, but in fact *precisely* optimal. This conclusion is not immediately justified, however, because the system (9.1-7) is not completely controllable. Problem 9.1-4 seeks further clarification of this point, and thus our discussion of the way properties of optimal systems carry over to the new controller arrangement is concluded.

In the remainder of this section, we shall discuss briefly the equations of the plant-controller arrangement when a Luenberger observer is used. We restrict attention to single-output systems. We first observe that there is no loss of generality involved in describing the plant and controller by state-space equations with a special coordinate basis when we are seeking to com-

pute the overall transfer function matrix, or the qualitative stability properties of the arrangement. Therefore, we may assume that the $n \times n$ matrix F of the plant in (9.1-1) has the special form of Sec. 8.3,—i.e.,

$$F = \left[\begin{array}{cccc:c} & & F_e & & b \\ \hdashline 0 & 0 & \cdots & 0 & 1 \end{array}\right] \tag{9.1-16}$$

where F_e is an $(n-1) \times (n-1)$ matrix, which is the transpose of a companion matrix form, and possesses desired eigenvalues.

Also,

$$h' = [0 \quad 0 \quad \cdots \quad 0 \quad 1]. \tag{9.1-17}$$

As we know from Sec. 8.3, x_e is given from

$$x_e = \begin{bmatrix} \bar{x}_e \\ y \end{bmatrix} \tag{9.1-18}$$

with

$$\dot{\bar{x}}_e = F_e \bar{x}_e + \bar{b}y + \bar{G}u. \tag{9.1-19}$$

The overbar denotes deletion of the last entry or row, as the case may be.

The control law that is implemented is

$$\begin{aligned} u &= u_{\text{ext}} + K'x_e \\ &= u_{\text{ext}} + (\bar{K})'\bar{x}_e + \bar{\bar{K}}'y \end{aligned} \tag{9.1-20}$$

where $\bar{\bar{K}}'$ is the last column of K'.

The plant-controller arrangement is described by the four equations (9.1-1), (9.1-18), (9.1-19), (9.1-20), together with equation $y = h'x$. A state vector for the overall plant-controller arrangement is, therefore, provided by the $(2n-1)$ vector

$$\begin{bmatrix} x \\ \bar{x}_e \end{bmatrix}.$$

However, a more appropriate, but equally acceptable, choice for state vector proves to be

$$\begin{bmatrix} x \\ \bar{x} - \bar{x}_e \end{bmatrix}.$$

This results in the following equation:

$$\frac{d}{dt}\begin{bmatrix} x \\ \bar{x} - \bar{x}_e \end{bmatrix} = \begin{bmatrix} F + GK' & -G\bar{K}' \\ 0 & F_e \end{bmatrix}\begin{bmatrix} x \\ \bar{x} - \bar{x}_e \end{bmatrix} + \begin{bmatrix} G \\ 0 \end{bmatrix} u \tag{9.1-21}$$

together with

$$y = h'x. \tag{9.1-22}$$

These two equations then imply again that the transfer function matrix relating u to y is $h'[sI - (F + GK')]^{-1}G$, and that the zero-input response of

the plant-controller arrangement is determined by the eigenvalues of $F + GK'$ and of F_e. The only difference between the Luenberger estimator and that considered earlier is that the $n \times n$ matrix $F + K_eH'$ is replaced by the $(n-1) \times (n-1)$ matrix F_e. The carrying over of the ancillary properties associated with optimality continues as before.

We consider now a brief example. As in the earlier example, we consider the plant with transfer function $1/s(s+1)$, modeled by

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \quad y = [1 \quad 0]x,$$

and we suppose that the control law that has to be implemented is

$$u = u_{\text{ext}} - [1 \quad 1]x$$

to ensure optimality. Suppose, also, that the estimator pole is required to be at $s = -3$. Let us change the coordinate basis to achieve this end. Using the methods and notations of Sec. 8.3, we obtain

$$T_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix},$$

and with $x_2 = T_2T_1x$ (rather than x_2 denoting the second component of x!), the new state-space equations are

$$\dot{x}_2 = \begin{bmatrix} -3 & -6 \\ 1 & 2 \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u, \quad y = [0 \quad 1]x_2$$

with $u = u_{\text{ext}} - [1 \quad 3]x_2$. Accordingly, the estimator equation is

$$\dot{\bar{x}}_{2e} = -3\bar{x}_{2e} + u - 6y.$$

The plant-controller arrangement is shown in Fig. 9.1-7, with only the controller shown in detail.

References [1] through [3] discuss another class of controllers, which are shown in reference [4] to have the same general structure as the controller obtained by cascading a Luenberger estimator with a linear control law. More precisely, the controller is a linear system with inputs consisting of u and y, and output the feedback component of u. However, the controllers do not actually generate a state estimate x_e. Moreover, although in the use of a Luenberger estimator the closed-loop transfer function matrix has no additional poles introduced in it as a result of using state estimate feedback rather than state feedback (because the additional states are uncontrollable), the closed-loop transfer function matrix resulting in [1] through [3] does have a number of additional poles introduced. For a single-input, single-output plant, the number of poles is $2n - 1$, the sum of the plant dimension and controller dimension. Design of the controllers is somewhat akin to the design of Luenberger observers; however, minimization of a performance index of

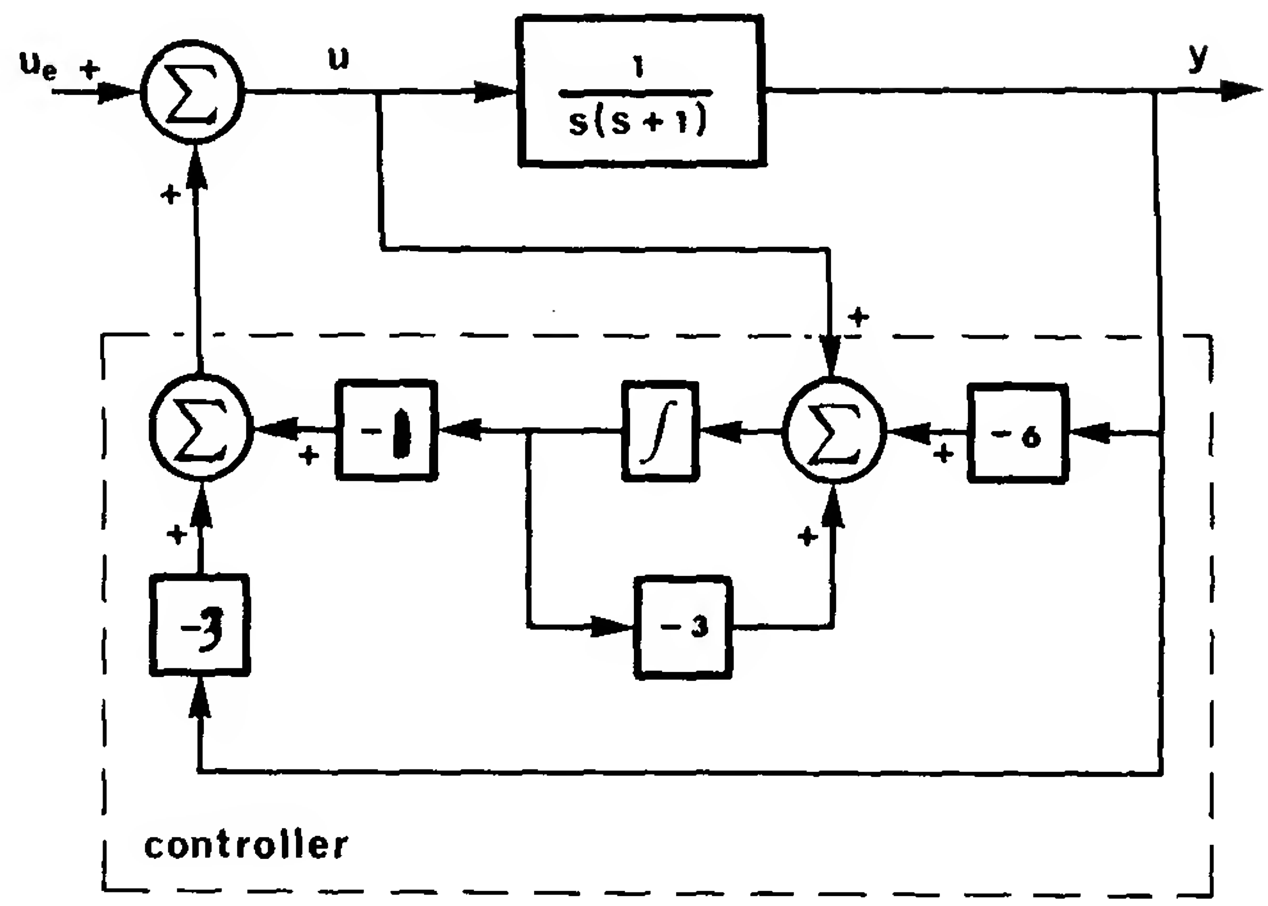


Fig. 9.1-7 Second example of a controller design.

the form $\int_{t_0}^{\infty} (w'Rw + [x'\ z']Q[x'\ z']')\,dt$ is required, where z is a vector containing various derivatives of u, and w is the pth derivative of u with $(p-1)$ the controller dimension. Some disadvantages of this form of controller over that already discussed are as follows:

1. Large nonlinearities cannot generally be tolerated at the plant input (even with adjustments to the controller input).
2. A higher order Riccati equation must be solved, which may present additional computational problems.
3. Estimates of the plant states are not available.
4. There is far less control over the additional modes introduced into the closed-loop system, and the closed-loop transfer function normally possesses more than n poles.

Chapter 10 discusses the basic problem of optimizing a performance index containing derivatives of the control (the use of such performance indices for controller design, as noted previously, being only one application). A problem is included in Chapter 10 illustrating controller design using the ideas of [1] through [3].

Problem 9.1-1. Consider the plant

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \quad y = [1 \quad 0]x.$$

Calculate an optimal control law that will minimize the performance index $\int_{t_0}^{\infty} (u^2 + 2x_1^2 + 3x_2^2)\,dt$, and design controllers of both dimension 2 and dimension 1 with poles that have real parts with significantly larger magnitude than the closed-loop system poles.

Problem 9.1-2. Discuss the gain and phase margin properties of optimally controlled systems where a state estimator is used.

Problem 9.1-3. Consider the first-order plant

$$\dot{x} = x + u.$$

The control law $u = -2x$ is an optimal law for this plant. Design state estimators of dimension 1 with poles at -1, -5, -10. Then sketch the response of $x(t)$ given $x(0) = 1$ for the following eight cases. In case 1, no estimator is used, and in cases 2 through 8 an estimator is used.

1. The feedback law is $u = -2x$.
2. Estimator pole is at -1 and $x_e(t_0) = 0$.
3. Estimator pole is at -1 and $x_e(t_0) = \frac{1}{2}$.
4. Estimator pole is at -1 and $x_e(t_0) = -1$.
5. Estimator pole is at -5 and $x_e(t_0) = 0$.
6. Estimator pole is at -10 and $x_e(t_0) = 0$.
7. Estimator pole is at -10 and $x_e(t_0) = \frac{1}{2}$.
8. Estimator pole is at -10 and $x_e(t_0) = -1$.

Comment.

Problem 9.1-4. Consider the $2n$-dimensional system defined by Eqs. (9.1-1) through (9.1-3), and let $u = K'x$ be an optimal control law resulting from minimization of a performance index $\int_{t_0}^{\infty} (u'u + x'Qx)\,dt$. Show that if Q is positive definite, there exists a positive definite $\hat{Q}$ such that $u = K'x_e$ is the optimal control for a performance index $\int_{t_0}^{\infty} [u'u + (x' \quad x' - x_e')\hat{Q}(x' \quad x' - x_e')']\,dt$. Indicate difficulties in extending this result to the case where Q is singular. (The conclusions of this problem are also valid when Luenberger estimators are used.) [*Hint:* Take as the state vector z, where $z' = (x' \quad x' - x_e')$.]

9.2 VARIANTS ON THE BASIC CONTROLLER

The aim of this section is to indicate some parallels with classical control ideas. This will be done by exhibiting some variants on the controller arrangements of the previous section, the structures of which will be familiar from classical control.

In general terms, we shall derive a controller structure of the form shown in Fig. 9.2-1. In classical nomenclature, there is a series or cascade compensator and a feedback compensator. In Fig. 9.2-2, there is shown a second controller structure that we shall develop. However, it is not always possible to implement this structure for reasons to be discussed. This is in contrast to the first structure, which may always be implemented.

We now list some general properties of the controller structures.

1. The controller structures will be derived by manipulations on the con-

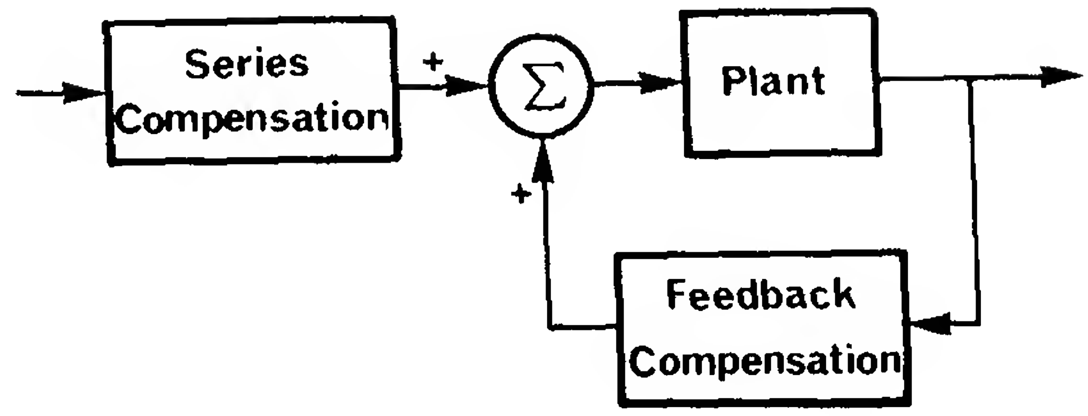


Fig. 9.2-1 Controller structure familiar from classical ideas.

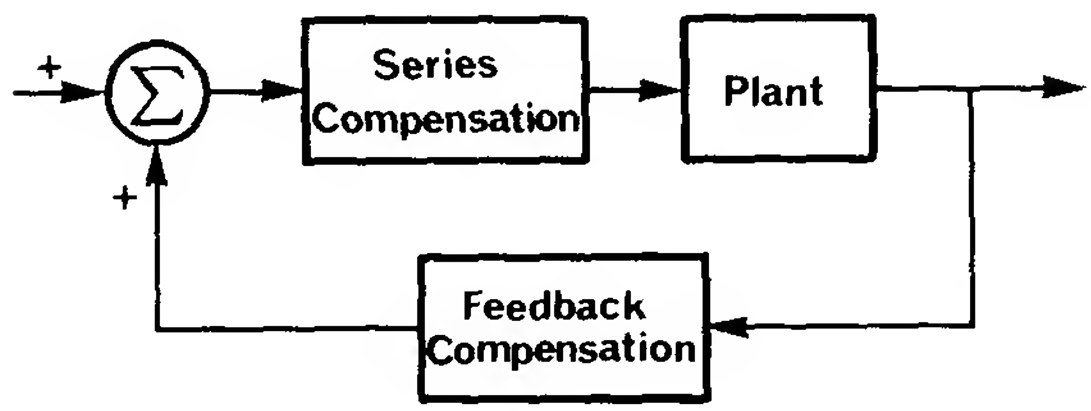


Fig. 9.2-2 Second controller structure familiar from classical ideas.

troller structure of the last section. These manipulations do not affect the input-output steady-state performance (or zero-state performance) of the overall scheme. In other words, the transfer function or transfer function matrix of the overall scheme is still $H'[sI - (F + GK')]^{-1}G$, where $H'(sI - F)^{-1}G$ is the transfer function (matrix) of the plant and $u = K'x$ is the desired feedback law.

2. There is no restriction to single-input or single-output plants, except, perhaps, an indirect one. Since the controller structures are derived from controllers of the sort described in the last section, one needs to be able to design versions of these for multiple-input, multiple-output plants. As we know, the multiple-output case may be hard.
3. The dimensions of the compensators are the same as the dimension of the controllers from which they are derived. This means that if, e.g., an n-dimensional single-output system employs an $(n - 1)$-dimensional controller, then the series compensator *and* the feedback compensator will each have dimension $(n - 1)$.
4. In view of 3, there are additional modes again in the overall scheme beyond those introduced in the controllers of the last section. In the case of the controller of Fig. 9.2-1, these additional modes are always asymptotically stable. However, this is not the case with the scheme of Fig. 9.2-2, which may thus prevent its use.

We shall now proceed with the derivation of the structures. We start with the controller of Fig. 9.2-3. Irrespective of the dimensionality of the controller, or its detailed design, there is a linear system between u and $K'x_e$, and a second linear system between y and $K'x_e$. In the steady state, the scheme of

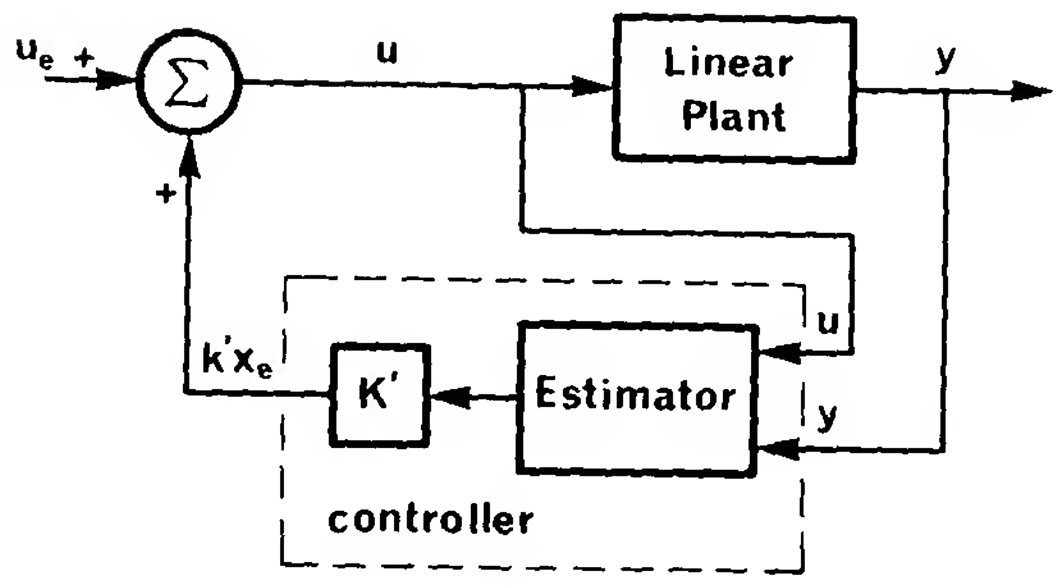


Fig. 9.2-3 The basic controller.

Fig. 9.2-3 is, in fact, equivalent to that of Fig. 9.2-4(a). In this figure, transfer function matrices are shown for each block. These are inserted to keep track of subsequent manipulations.

Figures 9.2-4(a), 9.2-4(b), and 9.2-4(c) are clearly equivalent. By setting $W_3 = [I - W_1]^{-1}$, Fig. 9.2-4(d) results. As will be shown, the inverse always exists. Figure 9.2-4(d), of course, has the same structure as Fig. 9.2-1. The equivalence between Figs. 9.2-4(d) and 9.2-4(e) follows by setting $W_4(s) = W_3(s)W_2(s)$. Figure 9.2-4(e) has the same structure as Fig. 9.2-2.

Thus, subject to establishing the existence of the inverse of $I - W_1$, the claims regarding the new controller structures have been established as far as their steady-state performance is concerned. What has yet to be checked is that the additional modes introduced have acceptable stability properties. To do this, and to check that the inverse exists, we now become specific. We start by considering estimators in Fig. 9.2-3 with the equation

$$\dot{x}_e = (F + K_e H')x_e + Gu - K_e y. \tag{9.2-1}$$

Of course, $F + K_e H'$ is assumed to have eigenvalues with negative real part. From (9.2-1), it follows immediately that the transfer function matrix relating u to $K'x_e$ is

$$W_1(s) = K'[sI - (F + K_e H')]^{-1}G \tag{9.2-2}$$

and the transfer function matrix relating y to $K'x_e$ is

$$W_2(s) = -K'[sI - (F + K_e H')]^{-1}K_e. \tag{9.2-3}$$

Forming $I - W_1(s)$, we can check the following formula for the inverse $[I - W_1(s)]^{-1}$ (which clearly establishes, at the same time, existence of the inverse):

$$W_3(s) = I + K'[sI - (F + K_e H' + GK')]^{-1}G. \tag{9.2-4}$$

Finally, the transfer function matrix $W_4(s)$, given by $W_3(s)W_2(s)$, can be computed to be

$$W_4(s) = -K'[sI - (F + K_e H' + GK')]^{-1}K_e. \tag{9.2-5}$$

[Problem 9.2-1 asks the student to verify the formulas for $W_3(s)$ and $W_4(s)$.]

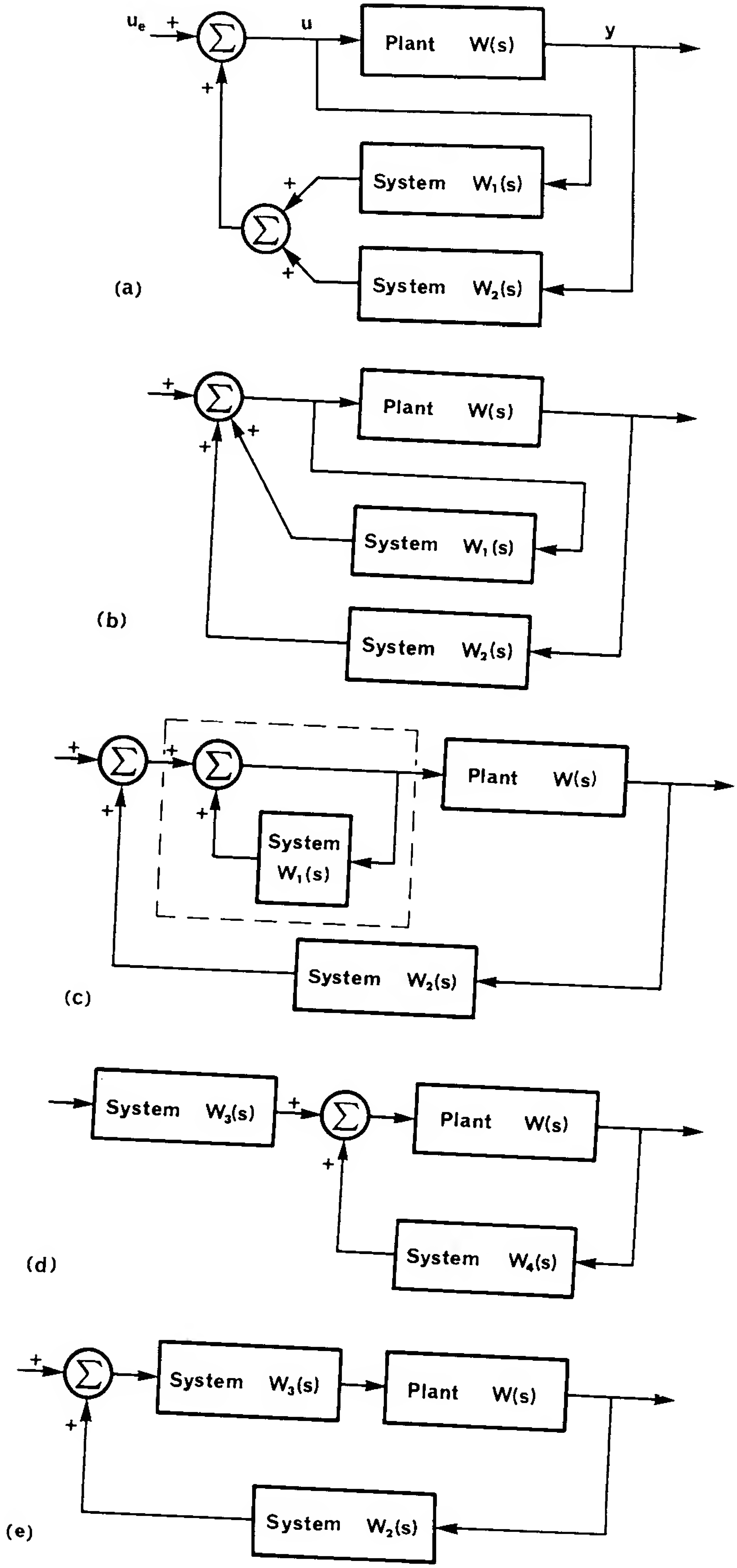


Fig. 9.2-4 Various system equivalences.

Let us now study the state-space equations of the scheme of Fig. 9.2-4(d), which will lead to conclusions regarding stability. We adopt the notation x_2 for the state of the linear system with transfer function $W_2(s)$, and x_3 for the state of the linear system with transfer function $W_3(s)$.

Now, the input to the $W_2(s)$ block is simply $H'x$. Therefore, this block may be described by [see (9.2-3)]:

$$\dot{x}_2 = (F + K_eH')x_2 - K_eH'x. \tag{9.2-6}$$

The output of the $W_2(s)$ block is, again from (9.2-3), $K'x_e$. From the block diagram it is then evident that the input to the $W_3(s)$ block is $u_{\text{ext}} + K'x_2$. Therefore, it is described by [see (9.2-4)]

$$\dot{x}_3 = (F + K_eH' + GK')x_3 + Gu_{\text{ext}} + GK'x_2. \tag{9.2-7}$$

The output of the $W_3(s)$ block is, again from (9.2-4), its input plus $K'x_3$, i.e., $u_{\text{ext}} + K'x_2 + K'x_3$. This is the input to the plant. Therefore, the plant equation is

$$\dot{x} = Fx + Gu_{\text{ext}} + GK'x_2 + GK'x_3. \tag{9.2-8}$$

Equations (9.2-6) through (9.2-8) constitute a set of state-space equations for the scheme of Fig. 9.2-4(d). An equivalent set is provided by retaining (9.2-6) and (9.2-8) and forming an equation for $\dot{x} - \dot{x}_2 - \dot{x}_3$. This equation follows at once from (9.2-6) through (9.2-8), and is

$$\dot{x} - \dot{x}_2 - \dot{x}_3 = (F + K_eH')(x - x_2 - x_3). \tag{9.2-9}$$

Equations (9.2-6), (9.2-8), and (9.2-9) combine to give

$$\begin{bmatrix} \dot{x}_2 \\ \dot{x} \\ \dot{x} - \dot{x}_2 - \dot{x}_3 \end{bmatrix} = \begin{bmatrix} F + K_eH' & -K_eH' & 0 \\ 0 & F + GK' & -GK' \\ 0 & 0 & F + K_eH' \end{bmatrix} \begin{bmatrix} x_2 \\ x \\ x - x_2 - x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ G \\ 0 \end{bmatrix} u_{\text{ext}} \tag{9.2-10}$$

The block triangular structure of the square matrix in (9.2-10) guarantees that the system modes are determined by the eigenvalues of $F + GK'$ and by those of $F + K_eH'$ counted twice. Consequently, the additional modes introduced beyond those of the basic controller of Fig. 9.2-3 (whose modes are determined by the eigenvalues of $F + GK'$ and $F + K_eH'$) are the eigenvalues of $F + K_eH'$. These modes are, of course, known to be asymptotically stable, and even under the designer's control. The validity of the proposed control arrangement is now established.

Two interesting points follow from Eq. (9.2-10). First, the last row of (9.2-10) [in reality, Eq. (9.2-9)] yields that $x_2 + x_3$ constitutes a state estimate

x_e of x. Second, setting

$$y = [0 \quad H' \quad 0]\begin{bmatrix} x_2 \\ x \\ x - x_2 - x_3 \end{bmatrix} \tag{9.2-11}$$

and adjoining this equation to (9.2-10), we can check from these two equations that the transfer function matrix relating u_e to y is actually $H'[sI - (F + GK')]^{-1}G$.

We now consider very briefly the situation that occurs if the control arrangement of Fig. 9.2-4(e) is used. Quite clearly, for the arrangement to be of any use, the transfer function matrix $W_4(s)$ must have the poles of all its elements in the half-plane $\text{Re}[s] < 0$, to ensure stability. This will be the case if and only if the matrix $F + K_eH' + GK'$ has eigenvalues all with negative real parts. Now, although the matrices $F + K_eH'$ and $F + GK'$ both have eigenvalues all with negative real parts, it does not follow at all that $F + K_eH' + GK'$ will have the same property. For example, if $F = [-1]$, $G = H = [1]$, and $K_e = K = [\frac{3}{4}]$, we have $F + K_eH' = F + GK' = [-\frac{1}{4}]$, whereas $F + K_eH' + GK' = [\frac{1}{2}]$.

Certainly, if $F + K_eH' + GK'$ does have negative real part eigenvalues, the scheme will be usable and have the correct steady-state behavior. In this case, the modes of the system can be shown to be determined by the eigenvalues of $F + GK'$, $F + K_eH'$, and $F + K_eH' + GK'$.

Calculations similar to those leading to $W_1(s)$, $W_2(s)$, etc., can be carried out when we start with a Luenberger estimator in Fig. 9.2-3. We shall assume that the state-space basis is so chosen that

$$F = \left[\begin{array}{ccccc|c} & & F_e & & & b \\ \hline 0 & 0 & \cdots & 0 & 1 & \end{array}\right] \tag{9.2-12}$$

where with F an $n \times n$ matrix, the matrix F_e is $(n-1) \times (n-1)$, is the transpose of a companion matrix, and possesses desired eigenvalues. We suppose also that y is a scalar, and, for convenience, that u is a scalar. As usual, an overbar denotes omission of the last row of a vector or matrix. Finally, $y = h'x = [0 \quad 0 \quad \cdots \quad 0 \quad 1]x$.

The estimator equation is, from the last chapter,

$$\dot{\bar{x}}_e = F_e\bar{x}_e + \bar{b}y + \bar{g}u \tag{9.2-13}$$

Now the full state estimate x_e is related to $\bar{x}_e$ and y by

$$x_e = \begin{bmatrix} \bar{x}_e \\ y \end{bmatrix} \tag{9.2-14}$$

and we are required to construct $k'x_e$. Denoting the last entry of k' by $\bar{\bar{k}}$, it follows that we have to construct

$$\bar{k}'\bar{x}_e + \bar{\bar{k}}y.$$

The transfer function relating u to this quantity is, then,

$$w_1(s) = \bar{k}'(sI - F_e)^{-1}\bar{g}, \tag{9.2-15}$$

and the transfer function relating y to this quantity is

$$w_2(s) = \bar{\bar{k}} + \bar{k}'(sI - F_e)^{-1}\bar{b}. \tag{9.2-16}$$

The transfer function $w_3(s)$ [see Fig. 9.2-4(d)] is $[1 - w_1(s)]^{-1}$, or

$$w_3(s) = 1 + \bar{k}'[sI - (F_e + \bar{g}\bar{k}')]^{-1}\bar{g}. \tag{9.2-17}$$

We shall not bother with evaluating the transfer function $w_4(s)$ of Fig. 9.2-4(e). This transfer function may or may not have poles in $\text{Re}[s] < 0$, depending on the precise matrices and vectors involved in arriving at it.

The existence of $[1 - w_1(s)]^{-1}$ guarantees the existence of the general control scheme of Fig. 9.2-4(d), and with this existence established, it follows that the scheme will give the correct steady-state properties. However, it is necessary to check the stability properties, because the arrangement introduces $(n - 1)$ further modes than are present when the usual Luenberger estimator is used. Suffice it to say, however, that similar analysis to that used for the case when non-Luenberger estimators are the starting point for a design yields a similar result: The modes of the controller-estimator scheme are determined by the eigenvalues of $F + gk'$, and the eigenvalues of F_e, each counted twice. Of course, this latter matrix is normally chosen to have a desired set of eigenvalues.

As an example, we consider a second-order, position-controller plant, with transfer function $1/s^2$, and with closed-loop poles at $-\frac{1}{2} \pm j\sqrt{\frac{3}{2}}$. Using a Luenberger estimator as the basis for the controller design, with its single pole at $s = -\alpha$, we shall deduce controller designs of the sort shown in Figs. 9.2-4(d) and 9.2-4(e). We shall examine subsequently how to choose the constant α.

The appropriate F, g, etc., matrices turn out to be as listed. (They may be readily derived by techniques discussed earlier.)

$$F = \begin{bmatrix} -\alpha & -\alpha^2 \\ 1 & \alpha \end{bmatrix} \quad g = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad h' = [0 \quad 1]$$

$$k' = [-1 \quad -(1 + \alpha)] \quad \text{and} \quad F_e = -\alpha.$$

The transfer function $w_2(s)$ of Eq. (9.2-16), which is the transfer function of the feedback compensator, is

$$w_2(s) = -(1 + \alpha) + \frac{\alpha^2}{s + \alpha} = -\frac{(1 + \alpha)s + \alpha}{s + \alpha}.$$

The transfer function $w_3(s)$ of Eq. (9.2-17), which is the transfer function of the series compensator, is

$$w_3(s) = 1 - \frac{1}{s + (\alpha + 1)} = \frac{s + \alpha}{s + \alpha + 1}.$$

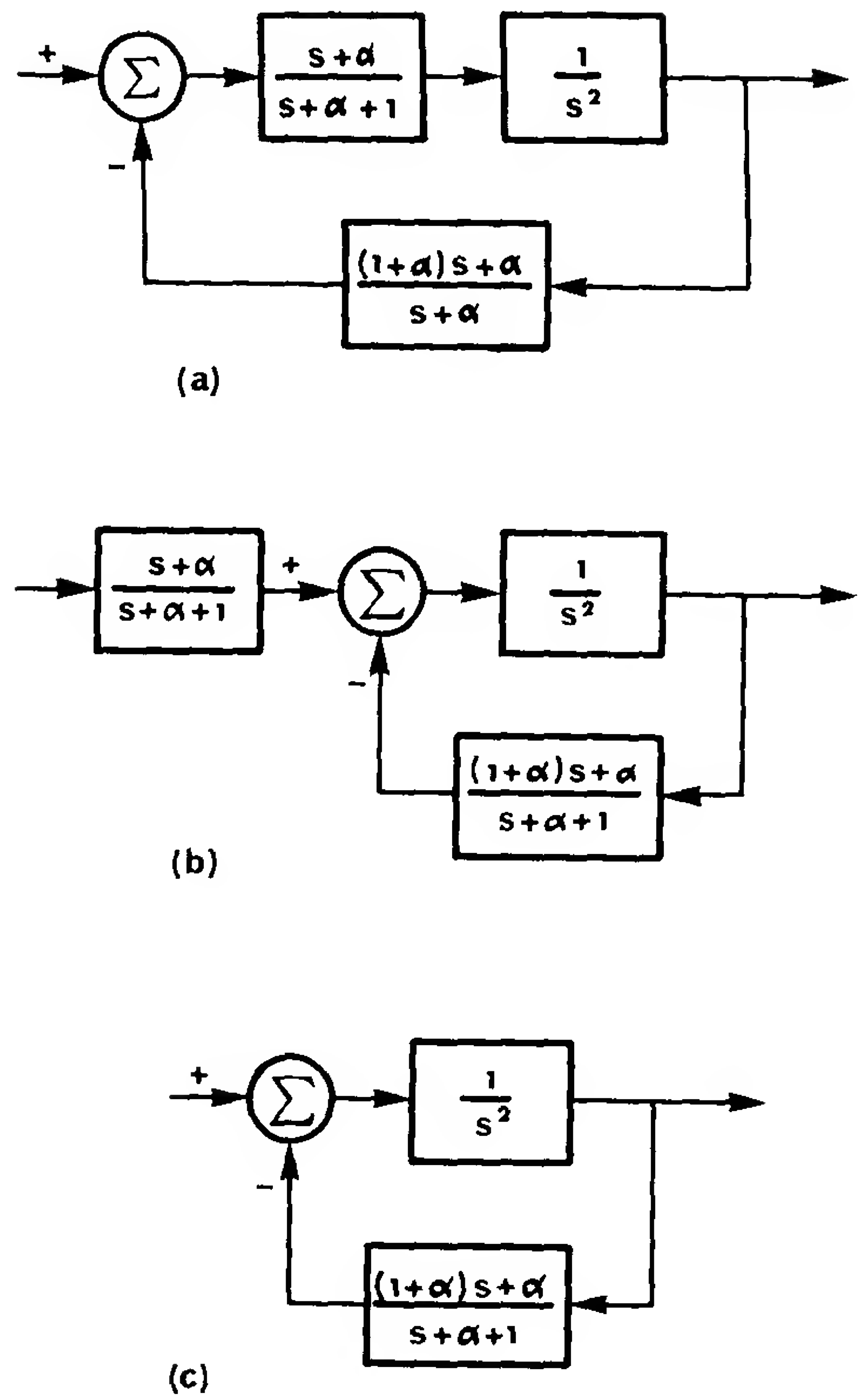


Fig. 9.2-5 Controllers discussed in example.

The control scheme is shown in Fig. 9.2-5(a). A scheme that is equivalent from the steady-state point of view is shown in Fig. 9.2-5(b).

With $\alpha = 10$, which would guarantee that the action of estimation was much more rapid than the action of the closed-loop plant, the feedback compensator transfer function becomes

$$\frac{11s + 10}{s + 10},$$

which is evidently a "lead network" transfer function, emphasizing high frequencies more than low frequencies. Evidently, the choice of larger α will increase further the extent to which high frequencies will be emphasized more than low frequencies, and thus increase the deleterious effects of noise. An upper limit of 10 on α is probably appropriate.

Figure 9.2-5(b) shows an arrangement with equivalent steady-state performance. The characteristic modes associated with this arrangement are

also close to those of Fig. 9.2-5(a). In the second case, they are determined by the zeros of $(s^2 + s + 1)(s + \alpha)(s + \alpha + 1)$, and in the first case by the zeros of $(s^2 + s + 1)(s + \alpha)^2$.

A classical design might actually have led to the scheme of Fig. 9.2-5(b), or, more probably, that of Fig. 9.2-5(c), which for $\alpha = 10$ is almost equivalent. The root locus diagram of the arrangement of Fig. 9.2-5(c) with $\alpha = 10$ is shown in Fig. 9.2-6, and it is evident what stabilizing effect the lead network has.

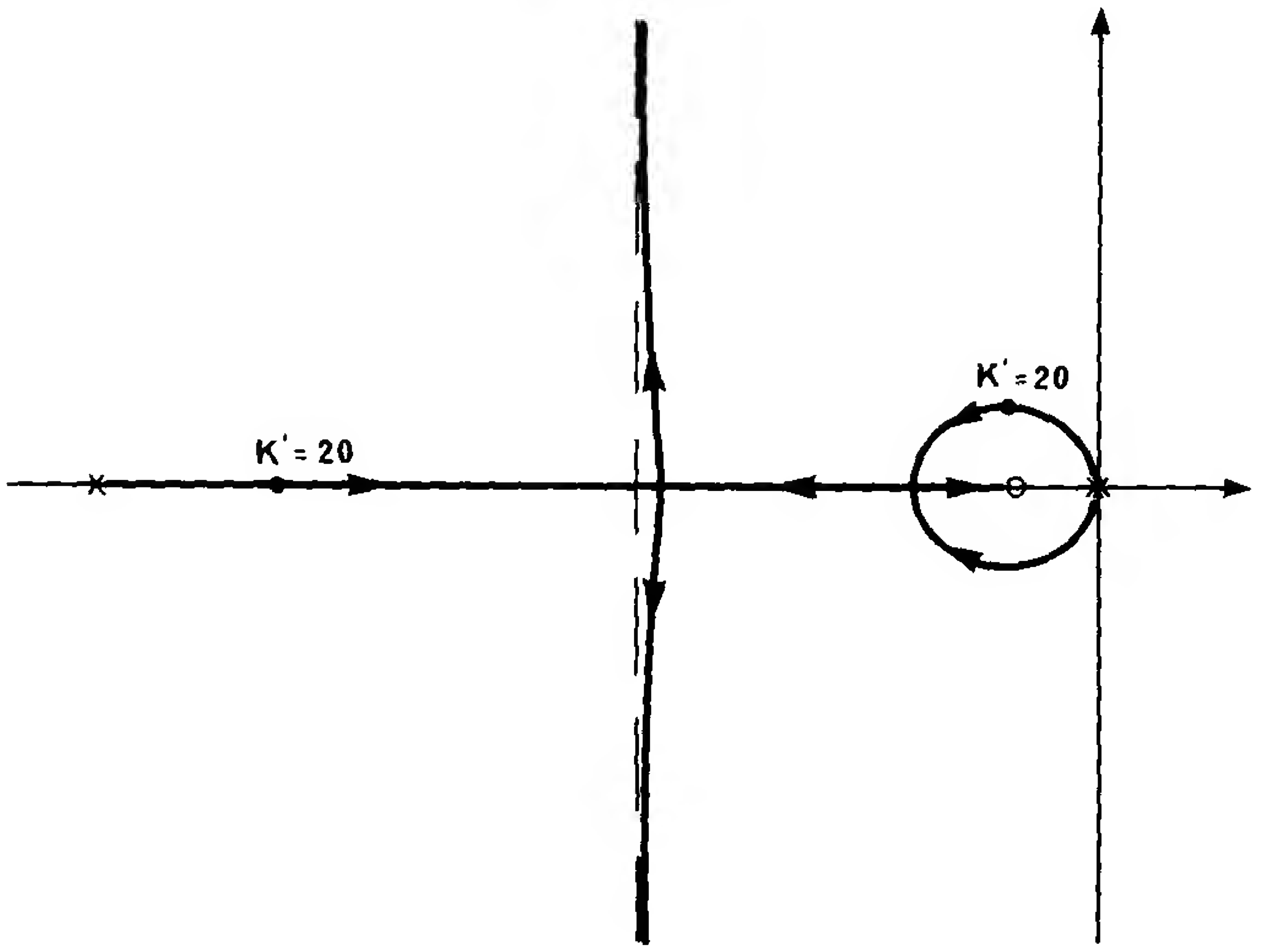


Fig. 9.2-6 Root locus plot for system of Fig. 9.2-5(c).

So far in this section, we have discussed the use of controllers of dimension n and $(n - 1)$ for controlling an n-dimensional, single-output plant. These controllers are constructed by cascading a linear control law with a state estimator. Now for special control laws, it may be that the controller dimension of n or $(n - 1)$ can be reduced. We give a qualitative and nonrigorous argument. It is intuitively reasonable that a single integrator subsystem driven by the plant input and output can generate an estimate of a single linear functional of the plant state. Now this linear functional, perhaps combined with the plant output, may be sufficient to generate a desired feedback signal, itself a linear functional; if this is the case, a controller of dimension 1 will then serve as well as a controller of dimension $(n - 1)$.

More generally, one can argue for multiple-input, multiple-output systems that certain special control laws can be implemented with controllers of dimension lower than usual. The characterization of such control laws and the design of the associated controllers is currently under investigation. It is interesting to speculate whether a linear optimal control problem may be

formulated for a prescribed plant, the solution of which will always lead to a low dimension controller.

In the next section, controller simplifications are discussed for multiple-output, single-input systems. These simplifications are well understood, but do not represent the best that can be achieved in specific situations.

Problem 9.2-1. Verify the formulas (9.2-4) and (9.2-5) for $W_3(s)$ and $W_4(s)$.

Problem 9.2-2. Design feedback and cascade compensators for a $1/s^2$ plant starting with a second-order estimator with estimator poles both at $s = -5$.

Problem 9.2-3. Consider a plant with transfer function $1/s^3$, and with closed-loop poles at $s = -\frac{1}{2} \pm j\sqrt{\frac{3}{2}}$ and $s = -3$. This corresponds to

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad g = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad h' = [1 \quad 0 \quad 0]$$
$$k' = [3 \quad 4 \quad 4].$$

Design feedback and cascade controllers based on a Luenberger estimator with poles at $s = -5$. Interpret the result using root locus ideas.

Problem 9.2-4. Obtain state-space equations for the controller arrangement of Fig. 9.2-4(d), assuming the controller blocks are determined from a Luenberger estimator. Check that the modes of the system are determined by the eigenvalues of $F + gk'$, and of F_e counted twice. [*Hint:* Obtain equations for x, x_2, and x_3, and then for $\bar{x} - x_1 - x_2$, x, and x_2, in that order, where $\bar{x}$ denotes x less its last entry.

9.3 SIMPLIFICATIONS OF THE CONTROLLER FOR MULTIPLE-OUTPUT, SINGLE-INPUT SYSTEMS

Although the remarks in this section will normally have extensions to the most general form of multiple-output system, we restrict consideration here almost entirely to multiple-output plants consisting of a cascade of single-input, single-output plants, each of whose outputs is available. Figure 9.3-1 shows a two-output example. The input of small plant 1 is the input of

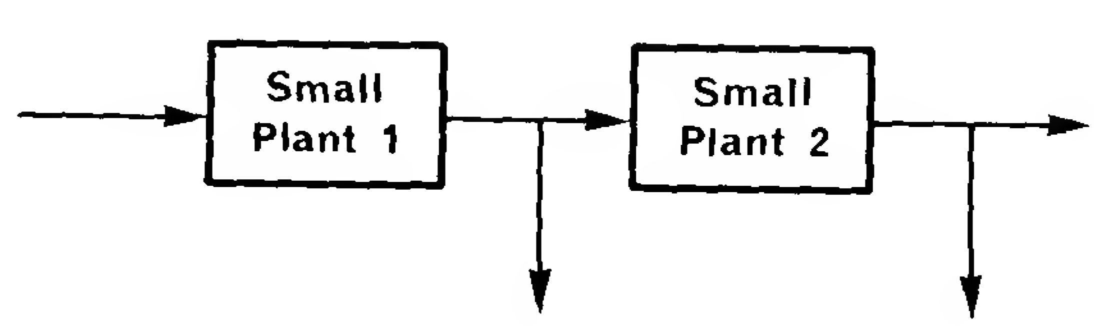


Fig. 9.3-1 Example of multiple-output plants considered.

the (large) plant, whereas a 2 vector, the entries of which are the outputs of small plants 1 and 2, is the (large) plant output.

We shall indicate in this section a procedure of [5] permitting a gross simplification in the estimator structure when a function of the form $k'x_e$ is required, rather than the complete vector x_e. The nature of the simplification is that the dimension of the whole estimator becomes one less than the maximum dimension of any of the small plants! Our method of attack in deriving this simplification will be to start with the basic Luenberger estimator and carry out manipulations on it.

As we know, the problem of estimating the states of the complete plant immediately breaks down into independent problems requiring the estimation of the states of each small plant. With the notation x_{ie} to denote an estimate of the state vector x_i of the ith small plant, and with u_i and y_i denoting the input and output of the ith small plant, the ith estimator can be written

$$\dot{\bar{x}}_{ie} = F_{ie}\bar{x}_{ie} + \bar{b}_i y_i + \bar{g}_i u_i \tag{9.3-1}$$

and

$$x_{ie} = \begin{bmatrix} \bar{x}_{ie} \\ y_i \end{bmatrix}. \tag{9.3-2}$$

Here, of course, we are assuming that

$$\dot{x}_i = \left[\begin{array}{ccccc|c} & & F_{ie} & & & b_i \\ \hline 0 & 0 & \cdots & 0 & 1 & \end{array}\right] x_i + g_i u_i \tag{9.3-3}$$

$$y_i = [0 \quad 0 \quad \cdots \quad 0 \quad 1]x_i \tag{9.3-4}$$

and F_{ie} has companion matrix transpose form with arbitrary eigenvalues, as chosen by the designer. Implicit in these equations is the assumption that an appropriate coordinate basis is used to write the small plant and estimator equations.

Figure 9.3-2(a) shows a plant consisting of two small plants, with estimators of the preceding type, and with the implementation of the control law $u = u_{\text{ext}} + k'x_e$. Observe that connecting any of the u_i or y_i with the feedback quantity $k'x_e$ is a single-input, single-output linear system. Each of these systems, of course, has a transfer function, and the denominator of the transfer function connecting u_i or y_i to $k'x_e$ is $\Delta_i(s) = \det(sI - F_{ie})$. Accordingly, from the steady-state point of view, Fig. 9.3-2(a) is equivalent to Fig. 9.3-2(b), where $m_1(s)$, etc., denote polynomials.

Now recall that the matrices F_{ie} are transposes of companion matrices, the characteristic polynomials of which are at the designer's disposal. To achieve a single controller design with two or more small plants, we suppose that all those F_{ie} of the highest dimension occurring among the complete set of F_{ie} are taken as the same, and that the remaining F_{ie} of lower dimension are taken to have characteristic polynomials that are factors of the charac-

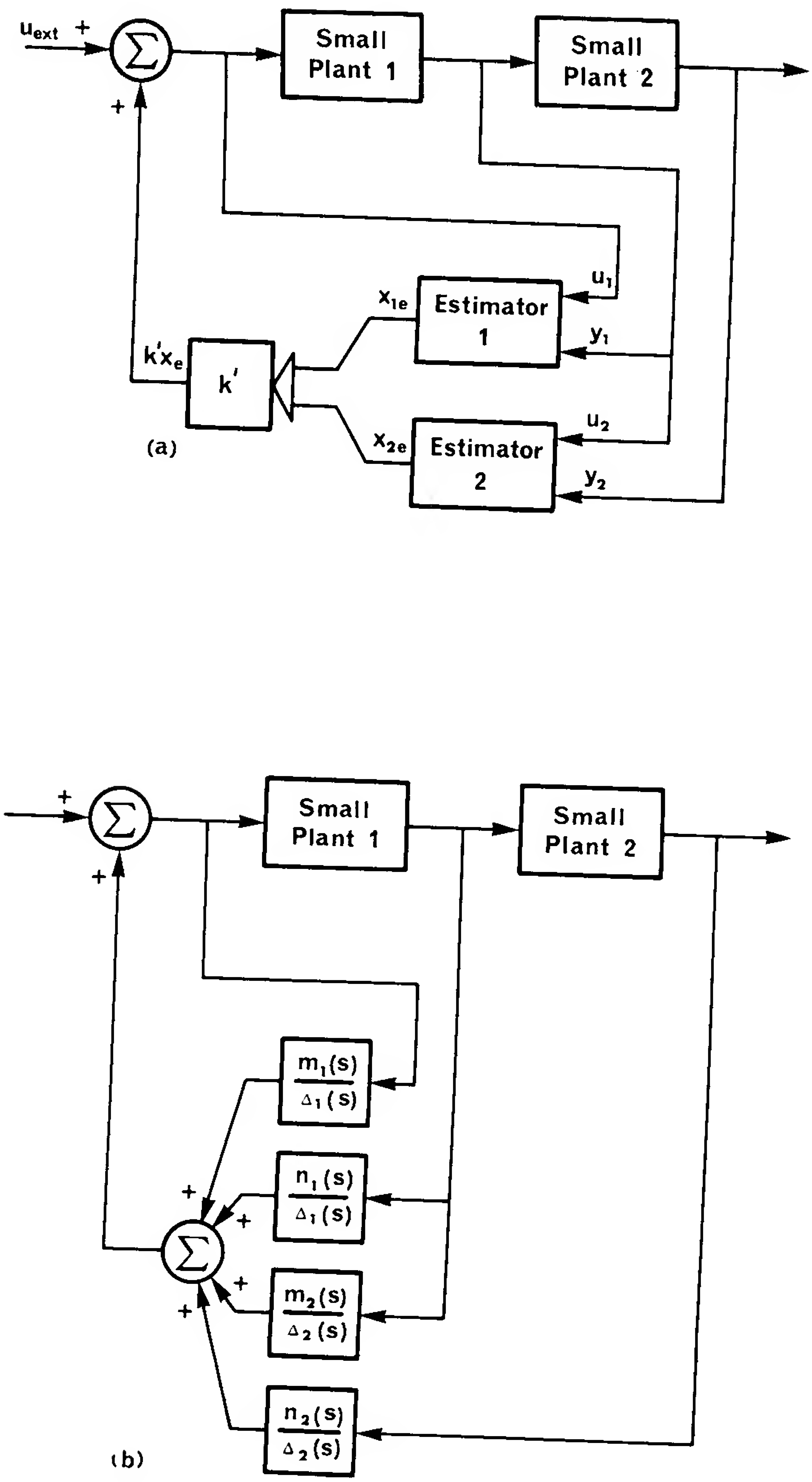


Fig. 9.3-2 Controller structure.

teristic polynomials of the F_{ie} of highest possible dimension. One or more of the $\Delta_i(s)$ will have maximum degree. Since the $\Delta_i(s)$ are adjustable, we require that all those of maximum degree be the same polynomial, call it $\Delta(s)$. All other $\Delta_i(s)$ must factor $\Delta(s)$. This means that any transfer function $m_i(s)/\Delta_i(s)$ may be written as $p_i(s)/\Delta(s)$, where $p_i(s) = m_i(s)$ if $\Delta_i(s)$ has the same degree as $\Delta(s)$, and $p_i(s) = m_i(s)\,[\Delta(s)/\Delta_i(s)]$ otherwise—i.e. $p_i(s)$ is obtained by multiplying $m_i(s)$ by an extra factor. The same thing is done with the transfer function $n_i(s)/\Delta_i(s)$, which may be written $q_i(s)/\Delta(s)$ for some polynomial $q_i(s)$.

Figure 9.3-3(a) shows the controller structure with these special choices, and with the augmentation of those transfer functions $m_i(s)/\Delta_i(s)$ and $n_i(s)/\Delta_i(s)$ by surplus numerator and denominator factors where necessary. Clearly, the blocks with transfer functions $q_1(s)/\Delta(s)$ and $p_2(s)/\Delta(s)$ [or, in the general case, $q_i(s)/\Delta(s)$ and $p_{i+1}(s)/\Delta(s)$], are immediately combinable by setting $r_2(s) = q_1(s) + p_2(s)$ [or, in the general case, $r_{i+1}(s) = q_i(s) + p_{i+1}(s)$]. With $r_1(s) = p_1(s)$ and $r_3(s) = q_2(s)$, this simplification results in Fig. 9.3-3(b).

We reiterate that the controllers of Fig. 9.3-2(a) and 9.3-2(b) are only equivalent in the steady state [since the controller of Fig. 9.3-2(b) has higher dimension than that of Fig. 9.3-2(a)]. The controller of Fig. 9.3-3(a) is obtained by specializing that of Fig. 9.3-2(b) and possibly by increasing further the dimension of the controller. It, too, is thus only equivalent in the steady state to the controller of Fig. 9.3-2(a). Finally, the controller of Fig. 9.3-3(b) is obtained by reducing the dimension of the controller of Fig. 9.3-3(a). However, it again is only equivalent to the controller of Fig. 9.3-2(a) (assuming appropriate selection of the matrices F_{ie}) in the steady state.

Now we can make one sweeping simplification to the arrangement of Fig. 9.3-3(b), resulting from the fact that the denominators of all the transfer functions are the same. It is possible to realize the arrangement of Fig. 9.3-3(b), again from the steady-state point of view, with a controller of dimension equal to the degree of $\Delta(s)$. One explicit way of writing down equations for such a controller is as follows. Each transfer function $r_i(s)/\Delta(s)$ is representable as noted in Appendix B by state-space equations of the form

$$\dot{w}_i = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -\alpha_1 \\ 1 & 0 & \cdot & \cdot & 0 & -\alpha_2 \\ 0 & 1 & \cdot & \cdot & 0 & -\alpha_3 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -\alpha_l \end{bmatrix} w_i + \begin{bmatrix} \gamma_{1i} \\ \gamma_{2i} \\ \gamma_{3i} \\ \cdot \\ \cdot \\ \gamma_{li} \end{bmatrix} v_i \tag{9.3-5}$$

$$z_i = [0 \quad 0 \quad \cdots \quad 0 \quad 1]w_i + \delta_i v_i \tag{9.3-6}$$

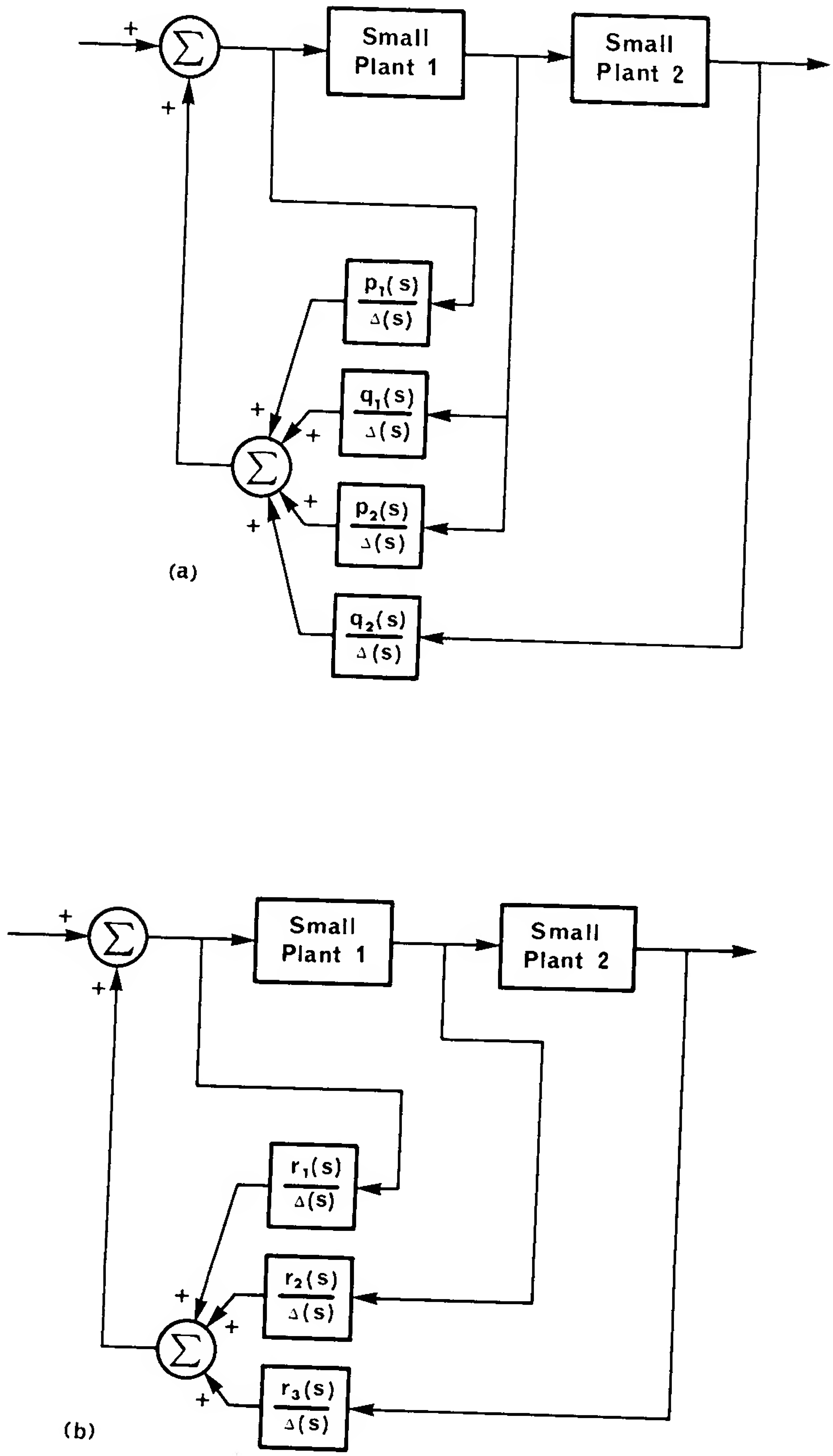


Fig. 9.3-3 Special cases of the controller of Fig. 9.3-2.

where v_i is the input into the block of transfer function $r_i(s)/\Delta(s)$, w_i is its state, and z_i its output. For the first block, v_1 coincides with the plant input u; for the second block, v_2 coincides with the output y_1 of small plant 1; for the third block, v_3 coincides with the output y_2 of small plant 2. The quantity fed back, which should be an estimate of $k'x$, is $\sum z_i$. (Of course, generalization to more than two small plants is immediate.)

The state-space equations of the overall controller become, simply,

$$\dot{w} = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & -\alpha_1 \\ 1 & 0 & \cdot & \cdot & 0 & -\alpha_2 \\ 0 & 1 & \cdot & \cdot & 0 & -\alpha_3 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 & -\alpha_l \end{bmatrix} w + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \gamma_{l1} & \gamma_{l2} & \gamma_{l3} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \tag{9.3-7}$$

$$z = [0 \quad 0 \quad \cdots \quad 1]w + [\delta_1 \quad \delta_2 \quad \delta_3] \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}. \tag{9.3-8}$$

Clearly, with the identification $z = \sum z_i$, the steady-state behavior of the controller defined by (9.3-5) and (9.3-6) is the same as that defined by (9.3-7) and (9.3-8). Again, generalization to many small plants is easy.

Figure 9.3-4 shows the new controller, where the matrices F_e and Γ of the figure have obvious interpretations. We reiterate that the dimension of F_e is one less than the maximum dimension of the small plants.

Although we have argued that only the steady-state performance of the controller is as required, it would appear that the extra modes that are introduced, above those which would be present in case true state feedback were employed, are all asymptotically stable. In fact, we claim that the modes of

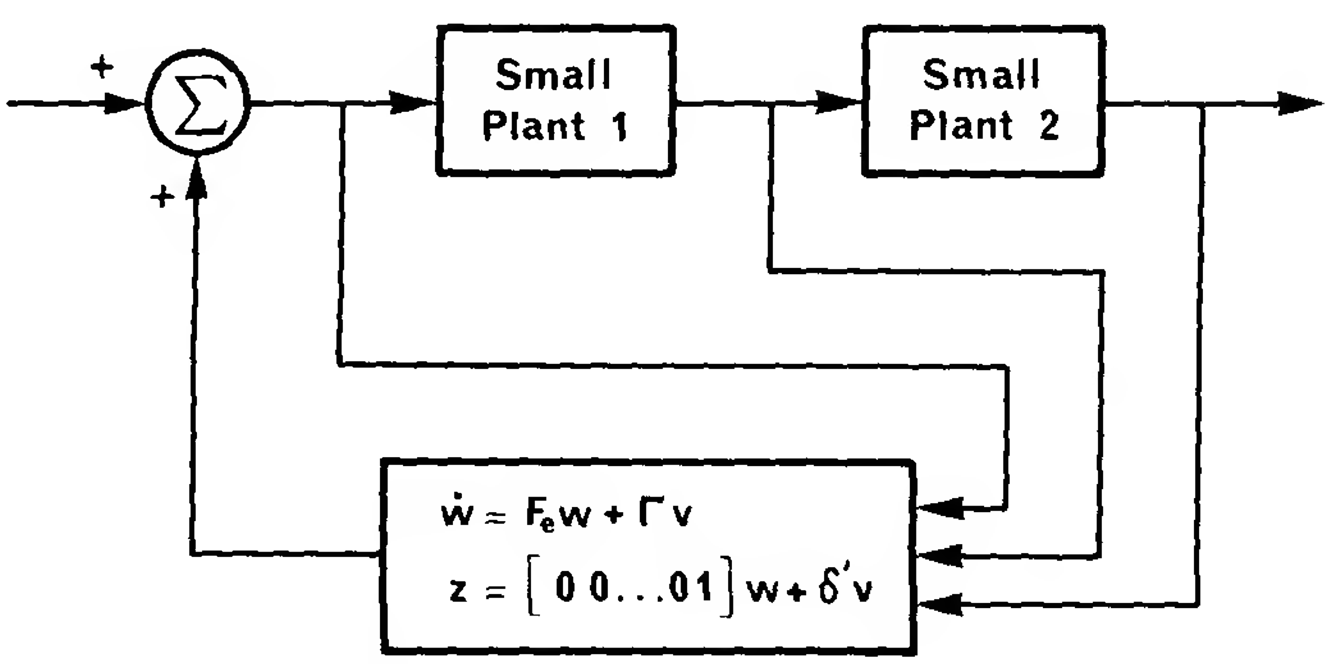


Fig. 9.3-4 Final controller arrangement.

the arrangement of Fig. 9.3-4 are determined by the eigenvalues of $F + gk'$ and the eigenvalues of F_e. However, we shall not prove this result here, which seems difficult to establish.

The preceding arguments will all hold if instead of using a Luenberger estimator for each small plant in Fig. 9.3-2(a), we use estimators whose dimensions are equal to the dimensions of the small plants. In this case, the end result is a controller with a dimension equal to the maximum dimension of any of the small plants.

We consider now a simple example. An open-loop plant is prescribed, which consists of a cascade of two elements—one with transfer function $1/s(s + 1)$ and the other with transfer function $2/s(s + 2)$ (see Fig. 9.3-5).

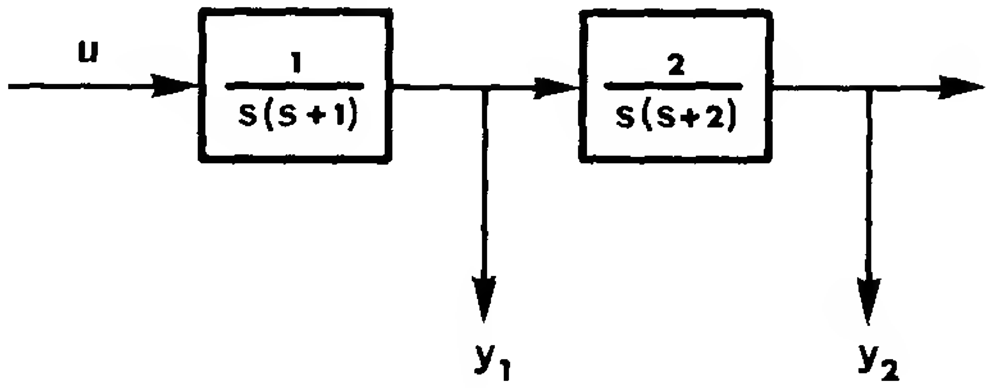


Fig. 9.3-5 Open-loop plant.

One state-space representation of the plant is

$$\dot{x}_1 = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

$$y_1 = [1 \quad 0]x_1$$

$$\dot{x}_2 = \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} x_2 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y_1$$

$$y_2 = [2 \quad 0]x_2.$$

Suppose that it is desired to implement the control law

$$u = u_{\text{ext}} + [-3 \quad -2 \quad -4 \quad -2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{9.3-9}$$

which, it can be checked, will yield closed-loop poles at $s = -\frac{1}{2} \pm j\sqrt{\frac{3}{2}}$ and a double pole at $s = -2$. A dynamic controller is to be used, the inputs of which are u, y_1, and y_2. It is to be of lowest possible dimension, here evidently 1, since the maximum dimension of the two small plants is 2.

The first task is to design estimators for the two small plants with the same dynamics. We shall locate the estimator poles at $s = -5$. With the procedures of the previous chapter, it is straightforward to establish that the scheme of Fig. 9.3-6(a) constitutes an estimator for the first small plant.

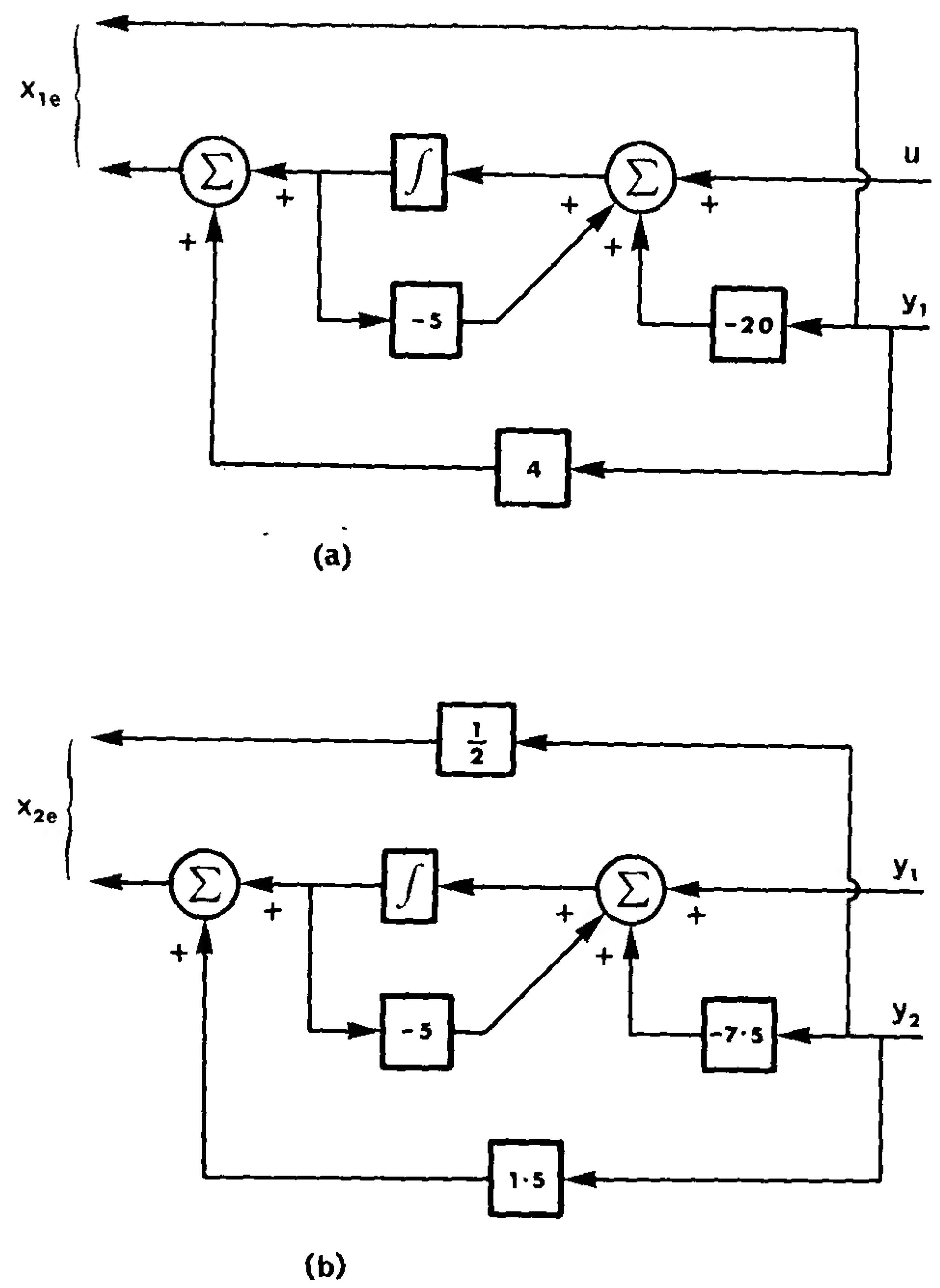
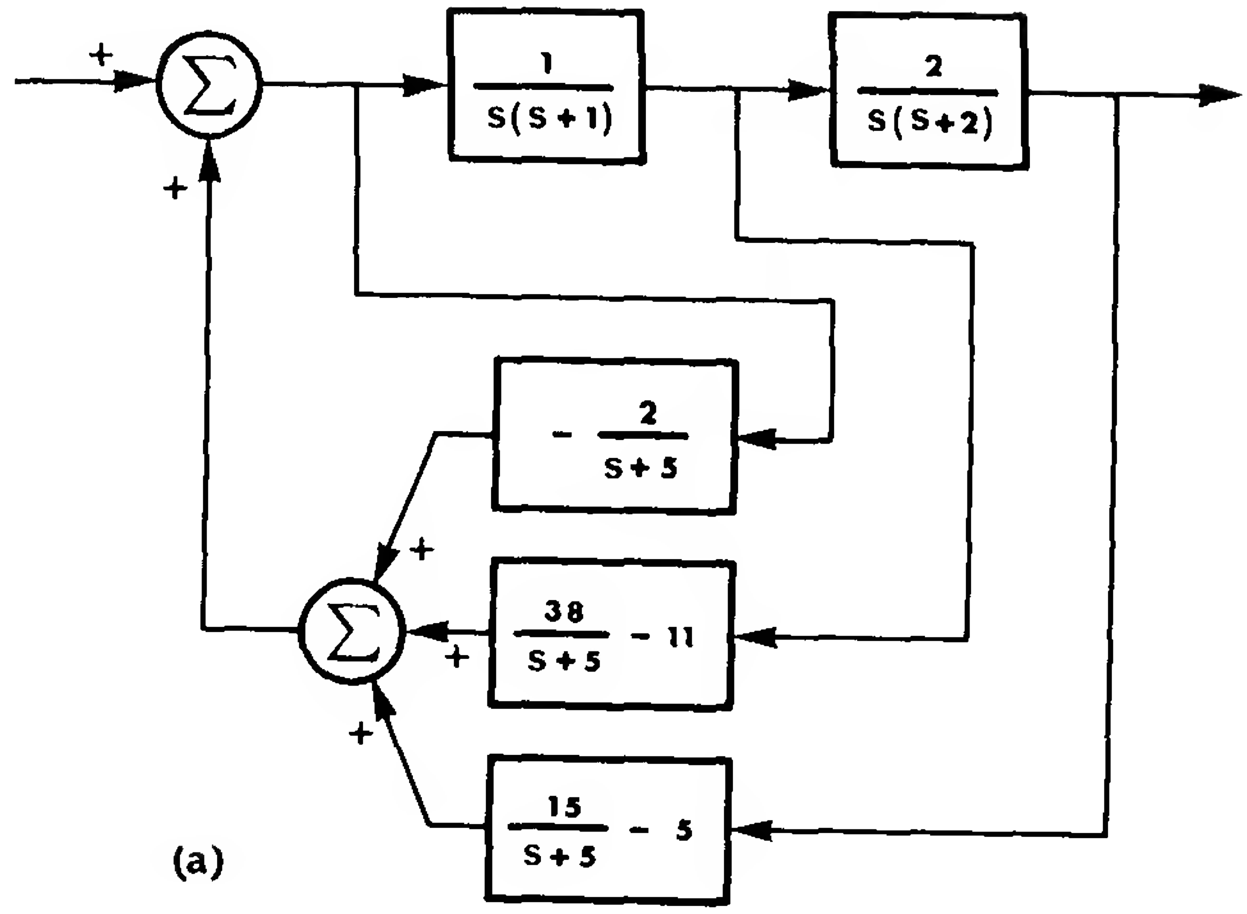


Fig. 9.3-6 Estimators for small plants.

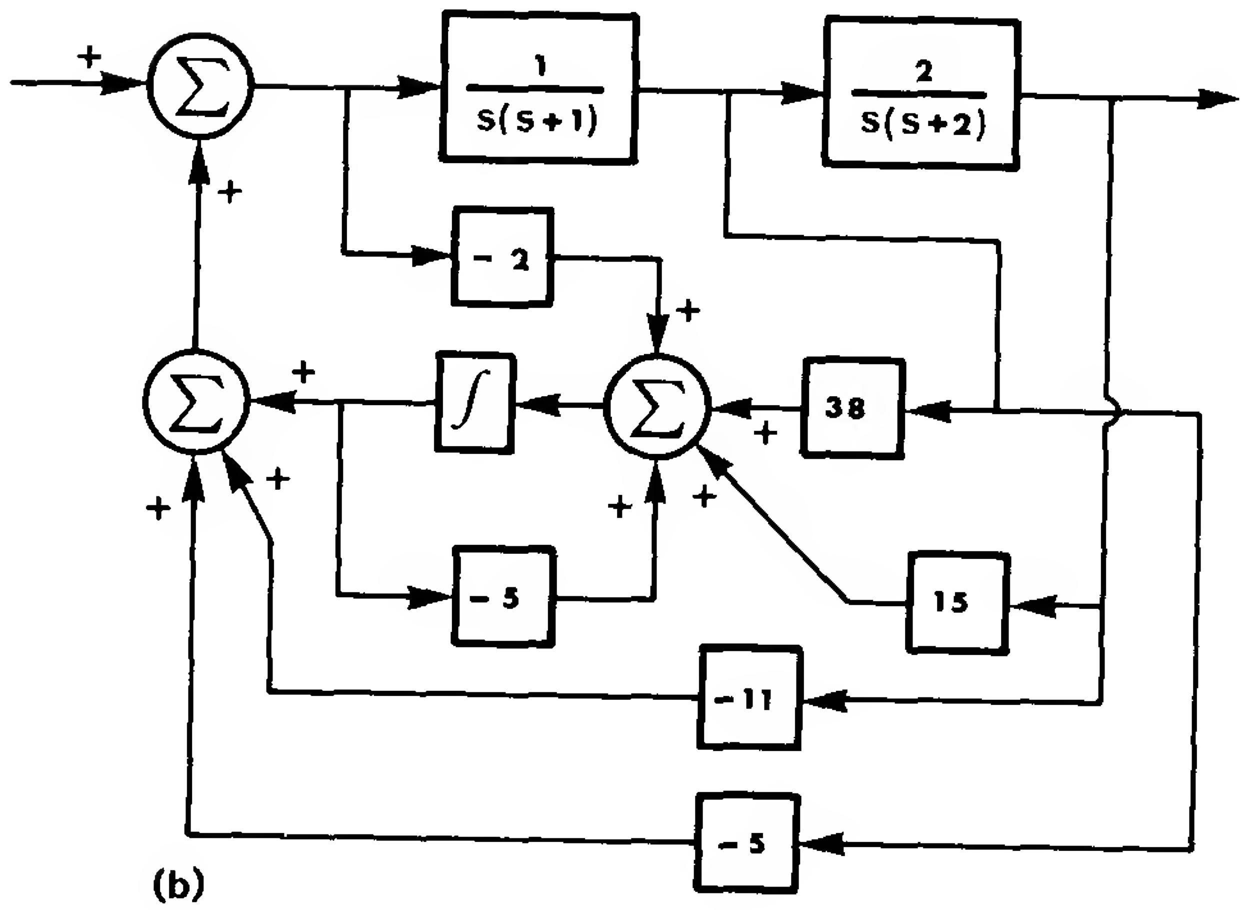
Likewise, the scheme of Fig. 9.3-6(b) constitutes an estimator for the second small plant.

When the control law (9.3-9) is implemented, we may proceed with the manipulation discussed earlier and shown in Figs. 9.3-2 and 9.3-3 to arrive at the control scheme of Fig. 9.3-7(a). Finally, the controller of Fig. 9.3-7(a) may be replaced by the scheme of Fig. 9.3-7(b), which is the desired controller of dimension unity.

If desired, the calculations can be checked most easily by separately checking the steady-state and zero-input response properties. Figure 9.3-7(a) is a convenient place from which to start when checking the steady-state response. A succession of equivalent systems (from the steady-state point of view) is shown in Fig. 9.3-8. The block manipulations involved are very straightforward, and the details are left to the reader to verify. Now the poles



(a)



(b)

Fig. 9.3-7

of the closed-loop system are at $-\frac{1}{2} \pm j\sqrt{\frac{3}{2}}$, with a double pole at -2, which means that the closed-loop system transfer function should be

$$\frac{2}{(s^2 + s + 1)(s + 2)^2}.$$

It is easy to check that $(s^2 + s + 1)(s + 2)^2(s + 5) = s^5 + 10s^4 + 34s^3 + 53s^2 + 44s + 20$, and so Fig. 9.3-8(d) does, in fact, represent the desired closed-loop system.

The zero-input response may be examined by writing down a state-space equation for the scheme of Fig. 9.3-7(b). With w denoting the state of

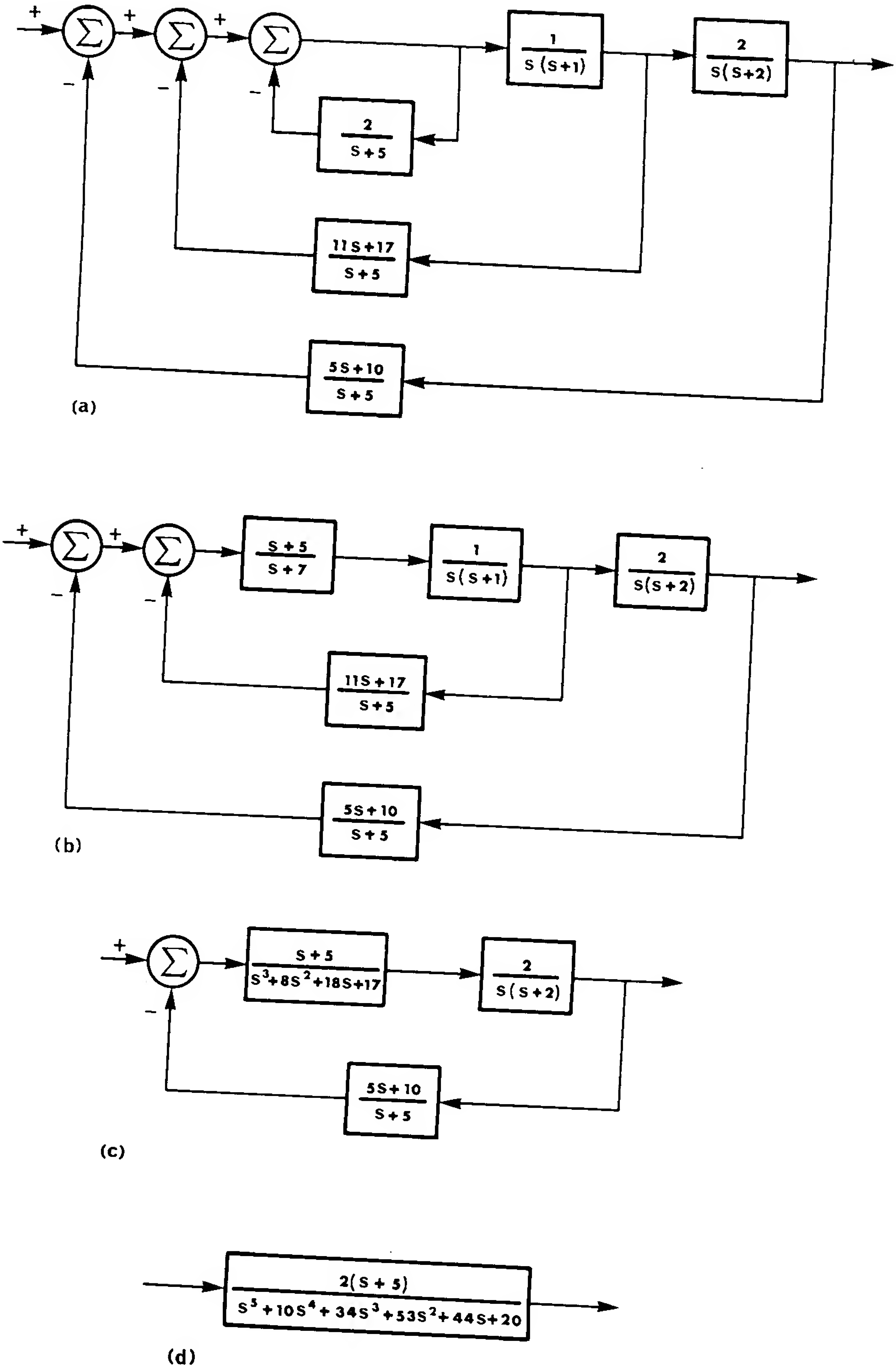


Fig. 9.3-8 Equivalent systems discussed in example.

the controller, and with $u_{ext} = 0$, this equation becomes

$$\begin{bmatrix} \dot{x}_1 \\ \\ \dot{x}_2 \\ \\ \dot{w} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -11 & -1 & -10 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -2 & 0 \\ 60 & 0 & 50 & 0 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ \\ x_2 \\ \\ w \end{bmatrix}$$

for which the associated characteristic polynomial may be verified to be $(s^2 + s + 1)(s + 2)^2(s + 5)$.

The extension of these results to multiple-output plants that do not consist of a cascade of single-input, single-output plants involves a good deal more computational effort [5]. The minimum dimension of the controller for a single-input, multiple-output plant is $\gamma - 1$, where γ is the *observability index* of the plant. The observability index is defined as follows. For a completely observable n-dimensional plant

$$\dot{x} = Fx + Gu \tag{9.3-10}$$

$$y = H'x, \tag{9.3-11}$$

the observability index γ is the smallest integer such that $[H \;\; F'H \;\cdots\; (F')^{\gamma-1}H]$ has rank n. If x is an n vector and y an m vector, γ evidently satisfies the inequality $n/m \leq \gamma \leq n - m$, provided that the entries of y are independent. Attainment of the upper bound means that there is no reduction in controller dimension below the dimension of the basic Luenberger estimator. Attainment of the lower bound, on the other hand, generally represents a substantial reduction in controller complexity.

The question of what is the minimum dimension of a controller for a multiple-input, multiple-output plant appears to be unanswered at the present time.

Problem 9.3-1. Consider a plant consisting of a cascade of small plants with transfer functions $1/(s + 1)$, $1/[s(s + 1)]$, and $1/s^2$. Obtain an optimal control law such that four dominant poles are located at $-\frac{1}{2} \pm j1$, -1, and -1. Design a controller of dimension 1 to implement this control law.

Problem 9.3-2. Can you give a general proof that the modes of the controller of Fig. 9.3-4 are determined by the eigenvalues of $F + gk'$ and F_e?

9.4 THE SEPARATION THEOREM

This section is confined to a brief statement of a theoretical result known as the "separation theorem" (see references [6] through [9]). We shall not attempt to prove the result, because the proof is not at all easy unless considerable background in statistics is assumed.

We assume we are given a linear system with additive input noise:

$$\dot{x} = F(t)x + G(t)u + v. \tag{9.4-1}$$

The input noise v is white, gaussian, of zero mean, and has covariance $\hat{Q}(t)\delta(t - \tau)$, where $\hat{Q}$ is nonnegative definite symmetric for all t. The output y of the system is given by

$$y = H'(t)x + w \tag{9.4-2}$$

where w is white gaussian noise of zero mean, and has covariance $\hat{R}(t)\delta(t - \tau)$, where $\hat{R}(t)$ is positive definite for all t. The processes v and w are independent. The initial state $x(t_0)$ at time t_0 is a gaussian random variable of mean m and covariance P_0, and is independent of the processes v and w. The matrices F, G, H, $\hat{Q}$, and $\hat{R}$ are all assumed to have continuous elements.

It is not possible to pose an optimal control problem requiring minimization of

$$V = \int_{t_0}^{t_1} (x'Q(t)x + u'R(t)u)\, dt \tag{9.4-3}$$

where $Q(t)$ is nonnegative definite symmetric and $R(t)$ is positive definite symmetric, even if one restricts the optimal $u(t)$ to being derived from the measurement $y(\cdot)$, because the performance index V must actually be a random variable, taking values depending on $v(\cdot)$, $w(\cdot)$, and $x(t_0)$, which, of course, are random.

To eliminate this difficulty, one can replace (9.4-3) by

$$V = E\left\{\int_{t_0}^{t_1} (x'Q(t)x + u'R(t)u)\, dt\right\} \tag{9.4-4}$$

where the expectation is over $x(t_0)$ and the processes $v(\cdot)$ and $w(\cdot)$ on the interval $[t_0, t_1]$. It is understood that at time t, the measurement $y(\tau)$, $t_0 \leq \tau \leq t$, is available, and that *the optimal $u(t)$ is to be expressed, in terms of $y(\tau)$, $t_0 \leq \tau \leq t$.* [*Note:* The optimal $u(t)$ is *not* required to be an instantaneous function of $y(t)$.]

The solution of this problem, which has come to be known as the separation theorem for reasons that will be obvious momentarily, is deceptively simple. It falls into two parts:

1. Compute a minimum variance estimate $x_e(t)$ of $x(t)$ at time t, using $u(\tau)$, $t_0 \leq \tau \leq t$ and $y(\tau)$, $t_0 \leq \tau \leq t$. As we know, this problem has a solution wherein $x_e(t)$ is the output of a linear system excited by $u(\cdot)$ and $y(\cdot)$. This linear system is independent of the matrices $Q(t)$ and $R(t)$—i.e., the same linear system generates $x_e(t)$, irrespective of what $Q(t)$ and $R(t)$ are.
2. Compute the optimal control law $u(t) = K'(t)x(t)$, which would be applied if there were no noise, if $x(t)$ were available, and if (9.4-3) were the performance index. Then use the control law $u(t) = K'(t)x_e(t)$, where $x_e(t)$ is obtained as in (1). This law is optimal for the

noisy problem. Notice that the calculation of $K(t)$ is independent of $H(t)$, and the statistics of the noise.

Evidently, the calculation of $x_e(t)$ and of the control law gain matrix $K(t)$ are separate problems which can be tackled independently, Hence, the name "separation theorem." Figure 9.4-1 shows the optimal controller.

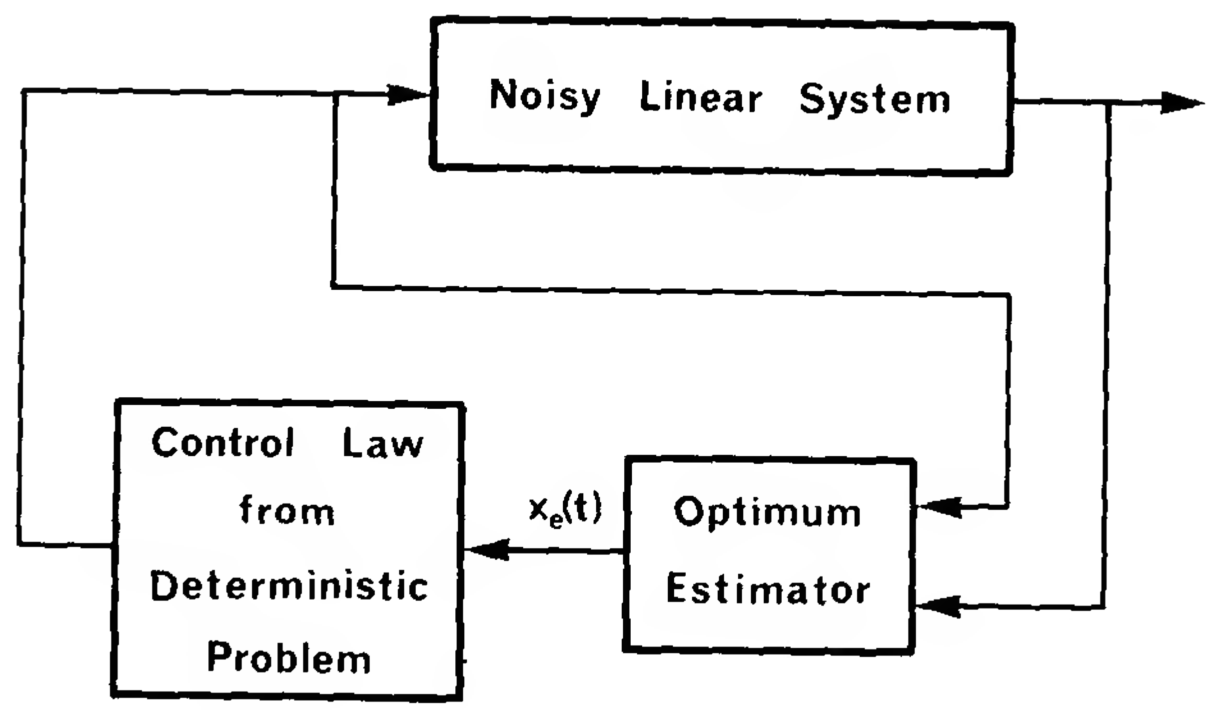


Fig. 9.4-1 Illustration of the separation theorem.

REFERENCES

[1] Pearson, J. B., "Compensator Design for Dynamic Optimization," *Intern. J. Control*, Vol. 9, No. 4, 1969, pp. 473–482.

[2] Pearson, J. B., and C. Y. Ding, "Compensator Design for Multivariable Linear Systems," *IEEE Trans. Auto. Control*, Vol. AC–14, No. 2, April 1969, pp. 130–134.

[3] Ferguson, J. D., and Z. V. Rekasius, "Optimal Linear Control Systems with Incomplete State Measurements," *IEEE Trans. Auto. Control*, Vol. AC–14, No. 2, April 1969, pp. 135–140.

[4] Moore, J. B., "A Note on Feedback Compensators in Optimal Linear Systems," *IEEE Trans. Auto. Control*, Vol. AC–15, No. 4, August 1970.

[5] Luenberger, D. G., "Observers for Multivariable Systems," *IEEE Trans. Auto. Control*, Vol. AC–11, No. 2, April 1966, pp. 190–197.

[6] Joseph, P. D., and J. T. Tou, "On Linear Control Theory," *Trans. Am. Inst. Elec. Eng.*, Pt. 2, Vol. 80, September 1961, pp. 193–196.

[7] Meier, L., "Combined Optimum Control and Estimation Theory," *Stanford Res. Inst. Contractor Rep. NASA*, Contract NAS 2-2457, October 1965.

[8] Striebel, C., "Sufficient Statistics on the Optimum Control of Stochastic Systems," *J. Math. Analysis Application*, Vol. 12, 1965, pp. 576–592.

[9] Wonham, W. M., "On the Separation Theorem of Stochastic Control," *SIAM J. Control*, Vol. 6, 1968, pp. 312–326.

PART IV

EXTENSIONS TO MORE COMPLEX PROBLEMS

CHAPTER 10

OPTIMAL LINEAR REGULATORS WITH INPUT DISTURBANCES

10.1 PROPORTIONAL-PLUS-INTEGRAL FEEDBACK

The linear optimal regulator theory of previous chapters is tailored for applications where the effect of an initial condition or, equivalently, the effect of impulse-type disturbances on a plant output is to be reduced to zero in some optimal fashion. However, for the case when there is a sustained, slowly varying, input disturbance, the standard optimal regulator cannot attain and maintain the desired equilibrium conditions.

A very reasonable approach to the linear regulator problem with input disturbances, including this latter type, would be to use a linear optimal regulator and to cancel the effect of the sustained input disturbance signal by the addition of a further control signal. This approach is certainly straightforward, provided there is a priori knowledge of the sustained disturbance, for all that is required is the addition of an appropriate, slowly varying, external input signal to cancel the effect of the disturbance. See, e.g., schemes given in reference [1]. Unfortunately, in most practical applications there is no a priori knowledge of the disturbance. The disturbance is probably randomly varying, as, e.g., in the case of an antenna control scheme subject to wind forces. In this instance, if the wind force is constant over an appreciable interval, or slowly varying over such an interval, there will usually be a steady-state error in the system output. That is, the point of equilibrium for the system will not be the desired operating point. To remedy the situation, some

mechanism making continual adjustments to the input is required. (In contrast, wind gusts are not a problem, because they consist of impulse disturbances with nonzero value over short intervals, and they can be accommodated by the standard regulator.)

Since the disturbances often encountered are randomly varying, it might be thought that optimal stochastic control theory (see, e.g., reference [2]) would be able to take into account the randomly varying disturbances. However, unless there is reliable a priori knowledge of the disturbance probability distributions, then this approach is also impractical.

The question is therefore still before us as to how to design a linear optimal regulator with the desirable properties that accrue from optimality, and with the additional property that sustained disturbances at the input of the type described do not affect the equilibrium point.

In classical control theory, the concept of *proportional-plus-integral feedback* is used to counter the effects of constant input disturbances, and *proportional-plus-multiple-integral feedback* is used to counter the effects of disturbances that can be expressed as a polynomial, e.g., a constant-plus-ramp input. Of course, with any design based on this theory, the resulting system does not necessarily have the desirable properties of optimal systems, and, unfortunately, the design procedures tend to be ad hoc, particularly for multiple-input systems. A review of some of the reasoning behind the concept of integral feedback is now given, since the concept will be applied in the next sections to linear optimal regulators.

For the case when the disturbance is a constant, the inclusion of a feedback path containing a single integration is sufficient for the closed-loop system equilibrium point to be zero, independent of the input disturbance. To see that this is so with a plausible but nonrigorous argument, we first suppose that the system output is zero. The output of the controller, in particular the integrator output, may still be a constant value. In fact, if it has a value precisely that of any constant input disturbance (but opposite in sign), the plant input will be zero, and thus the desired operating point will be maintained. On the other hand, if the system output is nonzero—say, a nonzero constant—then the output of the controller, and thus the input to the plant, will be a constant signal plus a ramp signal. The ramp signal will tend to reduce the output of the plant to zero (assuming that the feedback is of the appropriate sign) and thus tend to achieve the desired zero equilibrium point.

For the case when the input disturbance is a ramp function, as in the case of a velocity offset error, then the inclusion of a feedback path with two series integrations is used. For this case, the action of the controller once again is to provide a signal output when the system states are zero. This cancels the effects of the disturbances.

In this chapter, we construct two forms of a linear optimal regulator with dynamical feedback control laws. We then show that both systems have the various desirable properties associated with the standard regulator, and also the additional property that certain classes of input disturbances can be accommodated. (See also references [3] through [5].)

As a means to achieve the objectives of the chapter, the next section gives an extension of the standard regulator theory to include the case where the integrand of the performance index is quadratic, not only in the system states and in the system input but also in the derivatives of the system input. The optimal control is the solution of a differential equation involving the system states and inputs; thus, the controller for this case is dynamic rather than memoryless, as for the standard optimal regulator. Two ways of implementing the control law are then presented, which correspond to the proportional-plus-integral feedback of classical control.

The following section discusses the properties of this modified optimal regulator and its application to situations when there are unknown finite disturbances at the input to the plant to be controlled.

10.2 THE REGULATOR PROBLEM WITH DERIVATIVE CONSTRAINTS

We intimated in an earlier chapter that the standard regulator theory could be extended by the application of various transformations to give solutions to modified regulator problems. Using this approach, we shall solve the following problem.

Modified regulator problem. Suppose we are given the linear dynamical system

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{10.2-1}$$

where F and G are constant, and the performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (u'Ru + \dot{u}'S\dot{u} + x'Qx)\, dt \tag{10.2-2}$$

where S is positive definite symmetric, and R and Q are nonnegative definite symmetric. Suppose also that an initial value of the control $u(t_0)$ is specified. Find a control $u^*(\cdot)$ with $u^*(t_0) = u(t_0)$, which when applied to the system (10.2-1) minimizes (10.2-2). [Later, the case when $u(t_0)$ is to be optimized is considered.]

As a first step in solving this problem, we define new variables

$$x_1 = \begin{bmatrix} x \\ u \end{bmatrix} \quad u_1 = \dot{u} \tag{10.2-3}$$

and new matrices

$$F_1 = \begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix} \qquad G_1 = \begin{bmatrix} 0 \\ I \end{bmatrix} \qquad R_1 = S \qquad Q_1 = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}. \tag{10.2-4}$$

Equation (10.2-3) incorporates the transformations to be applied to the original system. Now, since $u(t_0)$ is specified, the system equations and performance index may be written in terms of the newly defined variables as follows:

$$\dot{x}_1 = F_1 x_1 + G_1 u_1 \qquad x_1(t_0) \text{ given} \tag{10.2-5}$$

$$V(x_1(t_0), u_1(\cdot), t_0) = \int_{t_0}^{\infty} (u_1' R_1 u_1 + x_1' Q_1 x_1)\, dt. \tag{10.2-6}$$

Figure 10.2-1 shows system (10.2-5); it consists of system (10.2-1) augmented by integrators at the inputs. The state vector of system (10.2-1) is augmented by the input vector of (10.2-1) to yield the state vector of system (10.2-5).

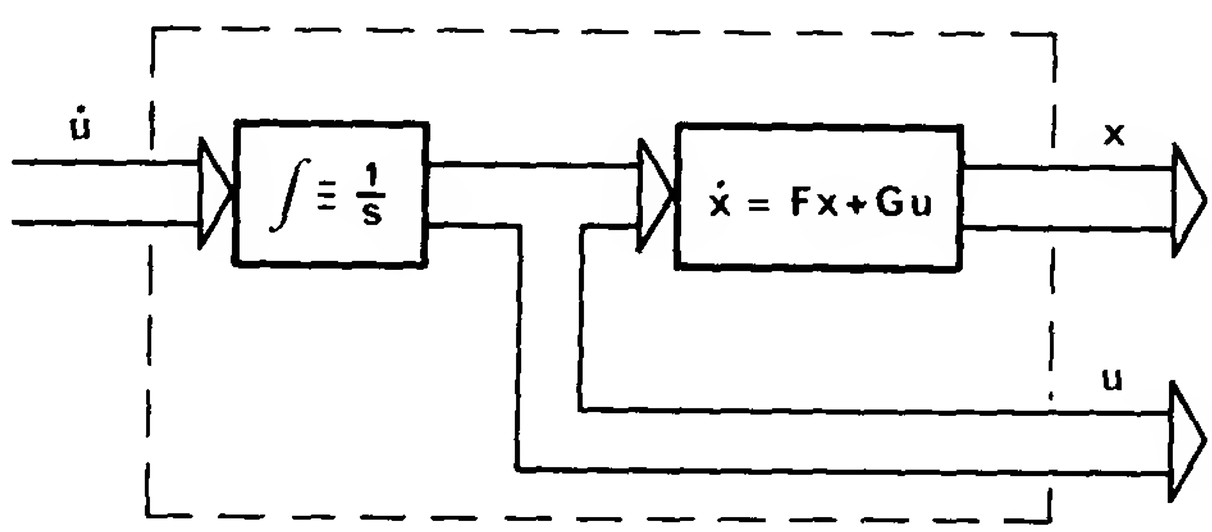


Fig. 10.2-1 Augmented system.

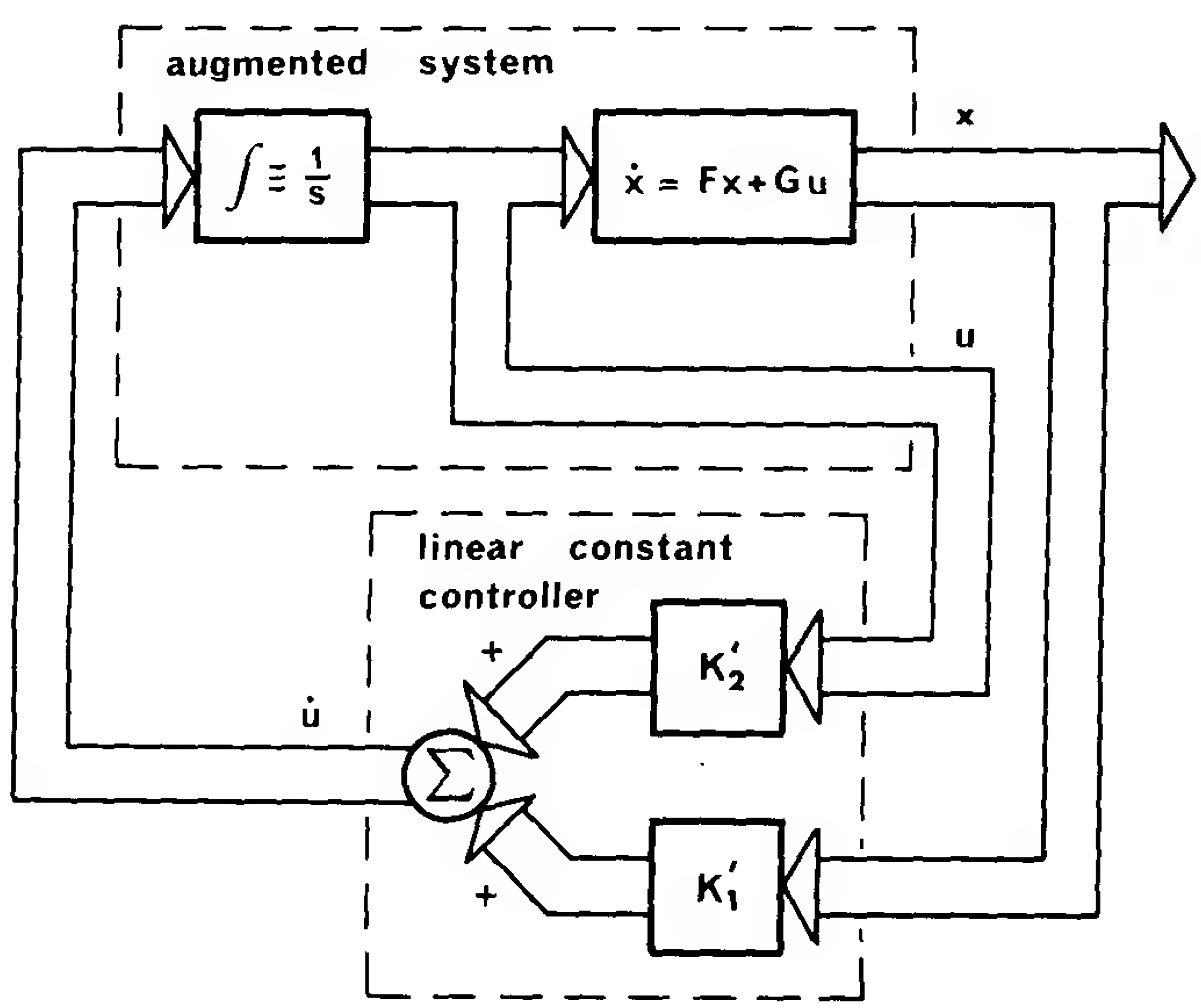


Fig. 10.2-2 Optimal control of augmented system.

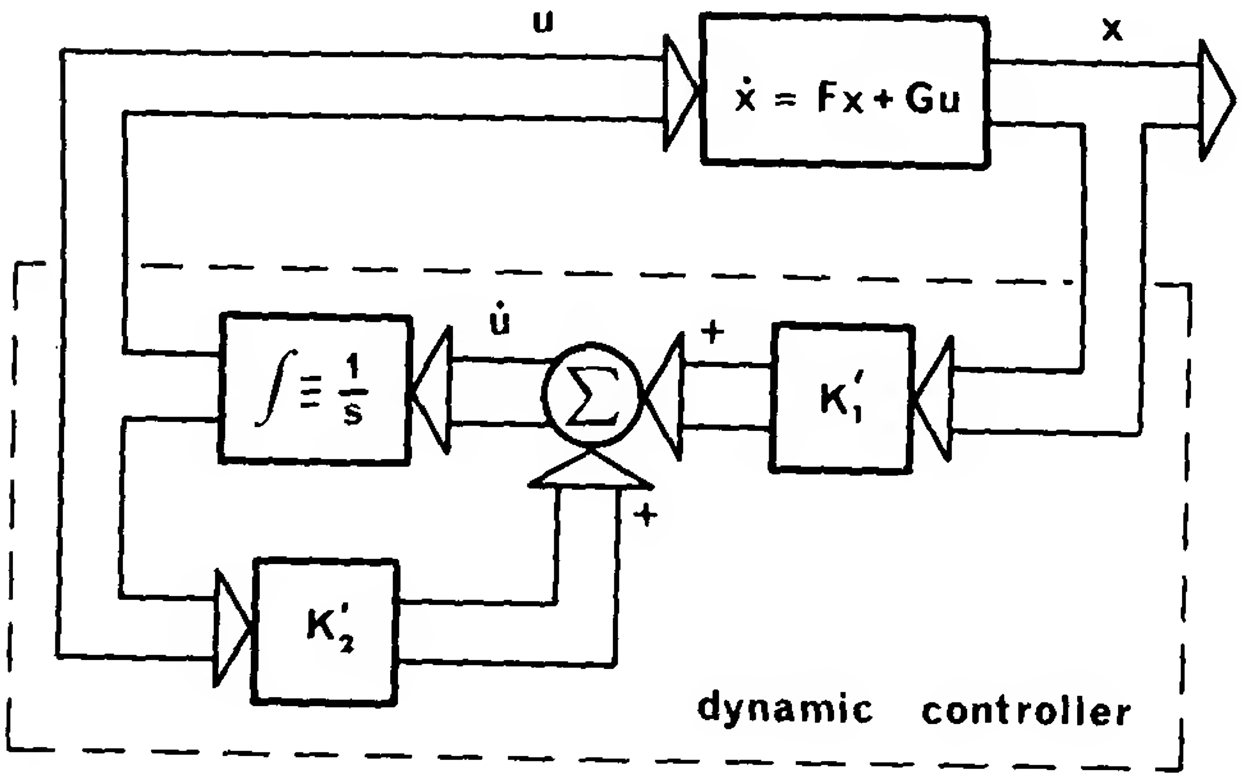


Fig. 10.2-3 Optimal regulator with dynamic controller.

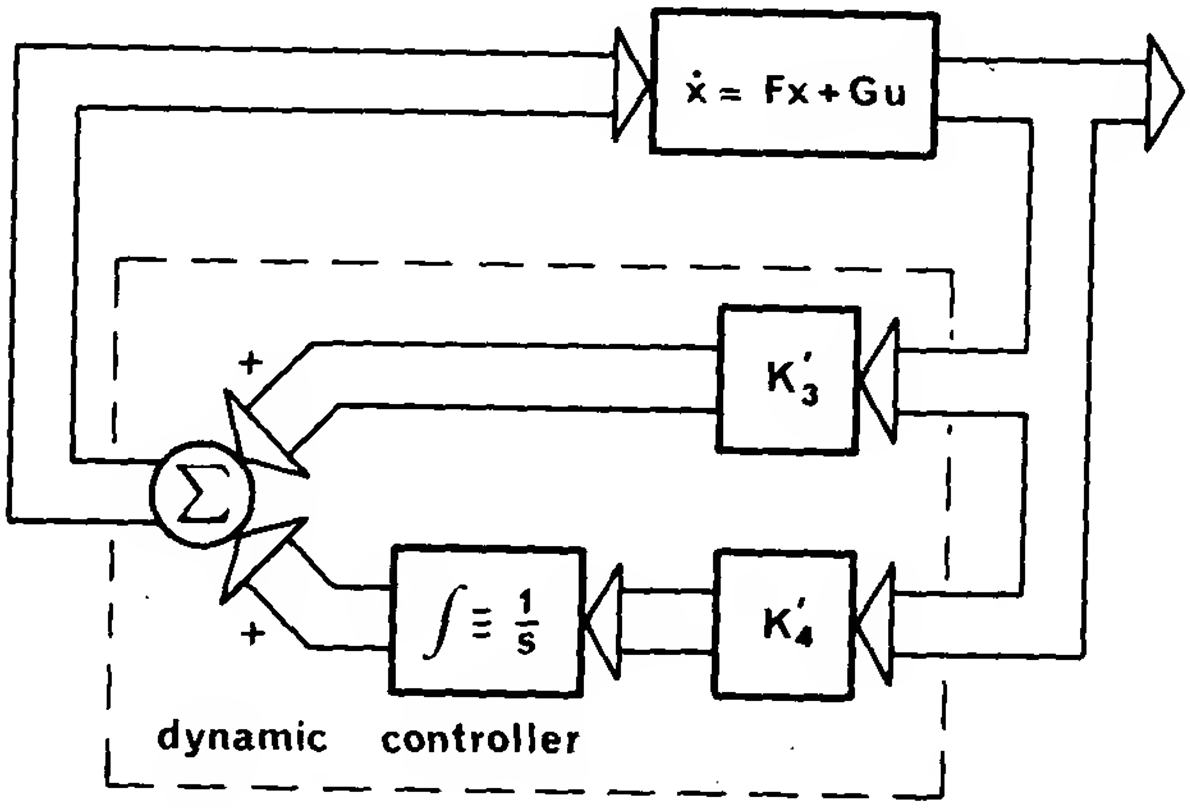


Fig. 10.2-4 Proportional-plus-integral state feedback.

The standard regulator theory of the earlier chapters may now be applied to the augmented system (10.2-5) and quadratic performance index (10.2-6). Of course, various controllability and observability conditions should by satisfied. Assuming that they are satisfied, the form of the optimal control is indicated in Fig. 10.2-2. This augmented system and linear memoryless controller may be rearranged as in Fig. 10.2-3 to be the original system (10.2-1) with a linear dynamic controller. In fact, later we shall show that an alternative arrangement is that of Fig. 10.2-4, where the controller has proportional-plus-integral state-variable feedback. That these regulators are, in fact, the desired optimal regulators is perhaps not immediately apparent, but this should become clear as we develop the subject in more detail.

First, we apply the standard regulator theory to minimize (10.2-6) subject to (10.2-5). We observe that the conditions previously imposed on

S, Q, and R ensure that R_1 and Q_1 satisfy the requirements of the standard regulator theory—viz., that R_1 be positive definite symmetric and Q_1 be nonnegative definite symmetric. However, two further assumptions are required in order that (1) the control law exist and the performance index be finite, and (2) the closed-loop system be asymptotically stable.

ASSUMPTION 10.2-1. The pair $[F_1, G_1]$ is completely controllable.

ASSUMPTION 10.2-2. The pair $[F_1, D_1]$ is completely observable for any D'_1 satisfying $D_1 D'_1 = Q_1$.

With the preceding assumptions, the optimal control law u_1^* and minimum index V^* associated with the minimization of (10.2-6), subject to (10.2-5), are given directly using the results of Chapter 3, Sec. 3.3. We have

$$u_1^* = -R_1^{-1}G'_1\bar{P}x_1 \tag{10.2-7}$$

and

$$V^*(x_1(t_0), t_0) = x'_1(t_0)\bar{P}x_1(t_0) \tag{10.2-8}$$

where

$$\bar{P} = \lim_{T\to\infty} P(t, T) = \lim_{t\to-\infty} P(t, T) \tag{10.2-9}$$

with $P(\cdot, T)$ the solution of the Riccati differential equation

$$-\dot{P} = PF_1 + F'_1P - PG_1R_1^{-1}G'_1P + Q_1 \quad P(T, T) = 0. \tag{10.2-10}$$

Assumption 10.2-1 ensures that $\bar{P}$ exists, and Assumption 10.2-2 ensures that the closed-loop system

$$\dot{x}_1 = (F_1 - G_1R_1^{-1}G'_1\bar{P})x_1 \tag{10.2-11}$$

is asymptotically stable.

We now examine, in order, the various previous equations, assumptions, and results in the light of the definitions (10.2-3) and (10.2-4). In this way, we shall obtain the solution of the modified regulator problem.

Earlier, we showed that the augmented system equation (10.2-5) follows from the original system equation (10.2-1), on using the definitions (10.2-3) and (10.2-4). The argument may be reversed without difficulty—i.e., (10.2-1) follows from (10.2-5), on using these same definitions. Furthermore, the definitions (10.2-3) and (10.2-4) imply that the two performance indices, (10.2-6) for the augmented system and (10.2-2) for the original system, take the same value. In particular, the optimal performance indices are the same when the initial states of the two systems are related by the first of Eqs. (10.2-3), and the optimal controls are related by the second of Eqs. (10.2-3).

We now ask if Assumptions 10.2-1 and 10.2-2 can be reduced to assumptions concerning the original system (10.2-1) and index (10.2-2). The answer is yes. In fact, Assumption 10.2-1 may be replaced by the following equivalent assumption.

ASSUMPTION 10.2-3. The pair $[F, G]$ is completely controllable.

Problem 10.2-1 asks for further details on this straightforward result.

In considering the observability Assumption 10.2-2, the arguments are more involved unless we make the further restriction that R be positive definite rather than simply nonnegative definite. For the simpler case, Assumption 10.2-2 may be replaced by Assumption 10.2-4.

ASSUMPTION 10.2-4. The matrix R is positive definite, and the pair $[F, D_{11}]$ is completely observable, where D_{11} is any matrix such that $D_{11}D'_{11} = Q$.

Problem 10.2-2 asks for details on this result. Notice that we have only claimed that Assumption 10.2-4 implies Assumption 10.2-2, and not the converse. The condition that R be positive definite may be relaxed with caution (see Problem 10.2-3).

Now that we have interpreted the equations and assumptions associated with the augmented system in terms of quantities associated with the original system, we turn to the interpretation of the results of the minimization problem for the augmented system, in order to give a solution to the minimization problem for the original system. As remarked earlier, the optimal control u^* for the modified regulator problem satisfies $\dot{u}^* = u_1$ with $u^*(t_0)$ equal to the specified $u(t_0)$. The minimum index, being the same for both minimization problems, is $V^*(x(t_0), u(t_0), t_0) = V^*(x_1(t_0), t_0)$. The optimal control u^* and the minimum index V^* can now be expressed in terms of the modified regulator problem parameters as follows.

Partitioning $\bar{P}$ as

$$\bar{P} = \begin{bmatrix} \bar{P}_{11} & \bar{P}'_{21} \\ \bar{P}_{21} & \bar{P}_{22} \end{bmatrix},$$

the optimal control u^* is given from

$$\begin{aligned} \dot{u}^* &= u_1^* \qquad u^*(t_0) = u(t_0) \text{ specified} \\ &= -R_1^{-1}G_1'\bar{P}x_1 \\ &= -S^{-1}\begin{bmatrix} 0 \\ I \end{bmatrix}' \begin{bmatrix} \bar{P}_{11} & \bar{P}'_{21} \\ \bar{P}_{21} & \bar{P}_{22} \end{bmatrix} \begin{bmatrix} x \\ u^* \end{bmatrix} \\ &= -S^{-1}\bar{P}_{21}x - S^{-1}\bar{P}_{22}u^*. \end{aligned}$$

That is, the optimal control $u^*(\cdot)$ is given from

$$\dot{u}^* = K_1'x + K_2'u^* \qquad u^*(t_0) = u(t_0) \text{ specified} \tag{10.2-12}$$

where

$$K_1' = -S^{-1}\bar{P}_{21}; \quad K_2' = -S^{-1}\bar{P}_{22}. \tag{10.2-13}$$

See Fig. 10.2-3 for a diagram of a controller that achieves this optimal con-

trol law. The minimum performance index associated with this control law is given as

$$\begin{aligned}V^*(x(t_0), u(t_0), t_0) &= V^*(x_1(t_0), t_0) \\ &= x_1'(t_0)\bar{P}x_1(t_0) \\ &= [x'(t_0) \quad u'(t_0)]\begin{bmatrix}\bar{P}_{11} & \bar{P}_{21}' \\ \bar{P}_{21} & \bar{P}_{22}\end{bmatrix}\begin{bmatrix}x(t_0) \\ u(t_0)\end{bmatrix} \\ &= x'(t_0)\bar{P}_{11}x(t_0) + 2u'(t_0)\bar{P}_{21}x(t_0) + u'(t_0)\bar{P}_{22}u(t_0).\end{aligned} \tag{10.2-14}$$

It is interesting to note that for the case when $u(t_0)$ is zero, the minimum index reduces to

$$V^*(x(t_0), u(t_0) = 0, t_0) = x'(t_0)\bar{P}_{11}x(t_0). \tag{10.2-15}$$

Until now, we have assumed that $u(t_0)$ is specified. However, it is clear that if $u(t_0)$ is chosen to minimize (10.2-14), then it would have the value

$$u^*(t_0) = -\bar{P}_{22}^{-1}\bar{P}_{21}x(t_0) = -(K_2')^{-1}K_1'x(t_0). \tag{10.2-16}$$

For this case,

$$V^*(x(t_0), t_0) = x'(t_0)[\bar{P}_{11} - \bar{P}_{21}'\bar{P}_{22}^{-1}\bar{P}_{21}]x(t_0). \tag{10.2-17}$$

These results are perhaps only of academic interest, because in practice there would usually be no more direct control of the initial states $u(t_0)$ of the controller than there would be of the initial states $x(t_0)$ of the plant.

The optimal control law and minimum index are expressed in terms of $\bar{P}_{11}$, $\bar{P}_{21}$, and $\bar{P}_{22}$. These matrices are simply the limiting values as t approaches minus infinity of P_{11}, P_{21}, and P_{22} where [see (10.2-9) and (10.2-10)]

$$\begin{aligned}-\dot{P}_{11} &= P_{11}F + F'P_{11} - P_{21}'S^{-1}P_{21} + Q \qquad & P_{11}(T) &= 0 \\ -\dot{P}_{21} &= P_{21}F + G'P_{11} - P_{22}S^{-1}P_{21} & P_{21}(T) &= 0 \\ -\dot{P}_{22} &= P_{21}G + G'P_{21}' - P_{22}S^{-1}P_{22} + R & P_{22}(T) &= 0.\end{aligned} \tag{10.2-18}$$

The results developed in the preceding paragraphs constitute a solution to the regulator problem with derivative constraints. However, there is an alternative formulation of the optimal control law u^*, which we now derive.

The input to system (10.2-1) may be expressed in terms of the state x and its derivative $\dot{x}$ from (10.2-1) as follows:

$$u = (G'G)^{-1}G'(\dot{x} - Fx). \tag{10.2-19}$$

For the preceding equation to be meaningful, we require that $G'G$ be positive definite or, equivalently, that G have rank equal to the number of system

inputs. For a completely controllable single-input system, $G'G$ will be a positive constant and thus invertible. For a multiple-input system, the physical interpretation of this condition on G is as follows. If the rank of G were not equal to the number of system inputs or number of columns of G, it would be possible to delete one or possibly more inputs and, by adjustment of the remaining inputs, to leave the system trajectories unaltered. In other words, the system is in a sense over controlled. Clearly, it is always possible to arrange the system such that this condition is avoided.

Assuming then that $G'G$ is positive definite, and thus that (10.2-19) holds, the differential equation (10.2-12) may be written

$$\dot{u}^* = K_3'\dot{x} + K_4'x \qquad u^*(t_0) = u(t_0) \text{ specified} \tag{10.2-20}$$

where

$$K_3' = K_2'(G'G)^{-1}G' \qquad K_4' = K_1' - K_3'F. \tag{10.2-21}$$

Integrating (10.2-20) gives

$$u^*(t) = K_3'x(t) + \int_{t_0}^{t} K_4'x(\tau)\,d\tau + u(t_0) - K_3'x(t_0). \tag{10.2-22}$$

From this equation it is clear that the optimal control $u^*(\cdot)$ can be realized by a proportional-plus-integral state feedback control law (see Fig. 10.2-4).

In practice, the initial state of the integrator in Fig. 10.2-4 will not be directly controlled any more than is the initial state of the plant $x(t_0)$. For the case in which we can choose an initial state for the integrator so that the performance index is minimized, clearly, from (10.2-16) and (10.2-22), this initial state would be $[(K_2')^{-1}K_1' - K_3']x(t_0)$.

We have the results that the optimal controller for our modified regulator problem may be either of the form shown in Fig. 10.2-3 or of the form shown in Fig. 10.2-4. The controllers for each case will be different but the resulting control $u^*(\cdot)$ will be the same. The equation of the closed-loop system in either case will be the same as that for the optimal augmented system of Fig. 10.2-2 —viz., (10.2-11) or, equivalently,

$$\begin{bmatrix} \dot{x} \\ \dot{u}^* \end{bmatrix} = \begin{bmatrix} F & G \\ K_1' & K_2' \end{bmatrix} \begin{bmatrix} x \\ u^* \end{bmatrix} \qquad \begin{bmatrix} x(t_0) \\ u^*(t_0) \end{bmatrix} \text{ specified.} \tag{10.2-23}$$

The various preceding results are now summarized.

The solution to the regulator problem with derivative constraints. For the completely controllable system (10.2-1) and performance index (10.2-2), the optimal control $u^*(\cdot)$ is given from

$$\dot{u}^* = K_1'x + K_2'u^* \qquad u^*(t_0) = u(t_0) \text{ specified} \tag{10.2-12}$$

or, equivalently, for the case when $(G'G)$ is positive definite, from

$$\dot{u}^* = K_3'\dot{x} + K_4'x \qquad u^*(t_0) = u(t_0) \text{ specified} \tag{10.2-20}$$

where K_1, K_2, K_3, and K_4 are given from (10.2-13), (10.2-18), and (10.2-21). (See also Figs. 10.2-3 and 10.2-4). The minimum index $V^*(x(t_0), u(t_0), t_0)$ is given as

$$V^*(x(t_0), u(t_0), t_0) = x'(t_0)\bar{P}_{11}x(t_0) + 2u'(t_0)\bar{P}_{21}x(t_0) + u'(t_0)\bar{P}_{22}u(t_0). \tag{10.2-14}$$

A sufficient condition for the closed-loop system

$$\begin{bmatrix} \dot{x} \\ \dot{u}^* \end{bmatrix} = \begin{bmatrix} F & G \\ K_1' & K_2' \end{bmatrix} \begin{bmatrix} x \\ u^* \end{bmatrix} \qquad \begin{bmatrix} x(t_0) \\ u(t_0) \end{bmatrix} \text{ specified} \tag{10.2-24}$$

to be asymptotically stable is Assumption 10.2-4. For the case when $u(t_0) = 0$, the minimum index is given as in (10.2-15), and for the case when $u(t_0)$ may be chosen so that the index is minimized, then the optimal $u(t_0)$ is given in (10.2-16) and the associated index in (10.2-17).

There is one immediate extension of the preceding results. By considering the augmented system (10.2-5) as the "original system" and applying the same procedure already outlined for this sytem, then we may interpret the results as constituting the solution of the regulator problem with second-order derivative constraints. The controllers for this case will contain a further stage of integration than those already discussed. The applications of this extension to the theory will be discussed in the next section, whereas the details of the extension process itself are left to the student (see Problem 10.2-4). Extensions are also possible to the case when the original system is augmented by an arbitrary dynamical system. These extensions have application (see references [6] through [8]) to the construction of dynamic compensators for systems with incomplete state measurements. Problem 10.2-5 asks for details to be worked out for a particular example.

Problem 10.2-1. Using the notation of the section, show that $[F_1, G_1]$ completely controllable is equivalent to the condition $[F, G]$ completely controllable. Use the controllability rank condition (see also Appendix B).

Problem 10.2-2. Using the notation of the section, show that Assumption 10.2-4 implies Assumption 10.2-2.

Problem 10.2-3. Suppose you are given the completely controllable system $\dot{x} = Fx + Gu$ and the performance index

$$V = \int_0^\infty (\dot{u}'S\dot{u} + x'Qx)\, dt.$$

Suppose also that D is a matrix satisfying $DD' = Q$, and that $[F, D]$ is completely observable. Show that the pair

$$\begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

fails to be completely observable if and only if there exist x_0 and u_0, not both zero, such that $Fx_0 + Gu_0 = 0$ and $D'x_0 = 0$. Give a simpler statement of this condition for the case when F is nonsingular. Show also that, if the condition holds, the closed-loop system is not asymptotically stable.

Problem 10.2-4. For the completely controllable system $\dot{x} = Fx + Gu$ and performance index

$$V = \int_0^\infty (\dot{u}\, S\dot{u} + x'Qx)\, dt,$$

find the optimal control law that minimizes the index V. Determine if the control law may be realized using proportional-plus-integral-plus-double-integral state-variable feedback.

Problem 10.2-5. For the system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

$$y = x_1 - x_2,$$

the control that minimizes the index

$$V = \int (\dot{u}^2 + x_1^2)\, dt$$

is

$$\dot{u} = -x_1 - 2x_2 - 2u.$$

Show how this control law may be realized using output, rather than state-variable feedback. Show that the feedback compensator has a structure similar to a Luenburger estimator with state-variable feedback. (*Hint:* First express $\dot{u}$ as

$$\dot{u} = -\alpha_0 u - \beta_0 y - \beta_1 \dot{y}.$$

See reference [6]. References [7] and [8] give generalizations of this procedure.)

10.3 APPLICATION AND PROPERTIES OF OPTIMAL REGULATORS WHICH INCLUDE INTEGRAL FEEDBACK

Using the theory of the previous section, linear optimal regulators that include single- or multiple-integral feedback may be designed in a straightforward manner. We now consider the properties of these systems, and, in particular, their ability to accommodate input disturbances.

Consider the completely controllable system with state equation

$$\dot{x} = Fx + G_1 u + G_2 \tilde{w} \qquad x(t_0) = 0, \quad u(t_0) = 0 \qquad (10.3\text{-}1)$$

where $\tilde{w}$ is a constant disturbance vector. Note that for simplicity we have chosen zero initial conditions. (Nonzero initial conditions can readily be

incorporated, if required.) Now, to achieve our objective—that the state x and its derivative $\dot{x}$ be zero as time approaches infinity—we clearly require the following relationship to hold:

$$G_1 u(\infty) = -G_2 \tilde{w}(\infty). \tag{10.3-2}$$

This condition can be satisfied if and only if it is possible to choose a matrix M such that $G_2 = G_1 M$, because then $u(\infty)$ can be chosen as $u(\infty) = -M\tilde{w}(\infty)$. That is, we require the following assumption.

ASSUMPTION 10.3-1. The range space of G_2 is contained in the range space of G_1.

With Assumption 10.3-1 holding, we may write the system equations (10.3-1) as

$$\dot{x} = Fx + G(u + w) \tag{10.3-3}$$

where $w = M\tilde{w}$, $G = G_1$, and $[F, G]$ is completely controllable.

For the system (10.3-3), a performance index with derivative constraints corresponding to that of the previous section, Eq. (10.2-2), is

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} [(u + w)'R(u + w) + \dot{u}'S\dot{u} + x'Qx]\, dt \tag{10.3-4}$$

where S is positive definite symmetric, and R and Q are nonnegative definite symmetric. Applying the results of the previous section, we have that the optimal control $u^*(\cdot)$ which minimizes the preceding index is given from either the differential equation

$$\dot{u}^* = K_1'x + K_2'(u^* + w) \qquad u^*(t_0) = 0 \tag{10.3-5}$$

or from the alternative equation

$$\dot{u}^* = K_3'\dot{x} + K_4'x \qquad u^*(t_0) = 0. \tag{10.3-6}$$

Note that we have used the fact that $(d/dt)(w + u) = \dot{u}$. The matrices K_1, K_2, K_3, and K_4 are given from

$$K_1' = -S^{-1}\bar{P}_{21} \quad K_2' = -S^{-1}\bar{P}_{22} \quad K_3' = K_2'(G'G)^{-1}G' \quad K_4' = K_1' + K_3'F. \tag{10.3-7}$$

The matrices $\bar{P}_{22}$ and $\bar{P}_{21}$ (and also $\bar{P}_{11}$) are given as the limiting values as t approaches minus infinity of P_{22} and P_{21} (and P_{11}). Here, P_{11}, P_{12} and P_{22} are given by

$$\begin{aligned}
-\dot{P}_{11} &= P_{11}F + F'P_{11} - P_{21}'S^{-1}P_{21} + Q & P_{11}(T) &= 0 \\
-\dot{P}_{21} &= P_{21}F + G'P_{11} - P_{22}S^{-1}P_{21} & P_{21}(T) &= 0 \\
-\dot{P}_{22} &= P_{21}G + G'P_{21}' - P_{22}S^{-1}P_{22} + R & P_{22}(T) &= 0.
\end{aligned} \tag{10.3-8}$$

The minimum index is also given for this minimization problem as

$$V^*(w, t_0) = w'\bar{P}_{22}w. \tag{10.3-9}$$

Furthermore, if the pair

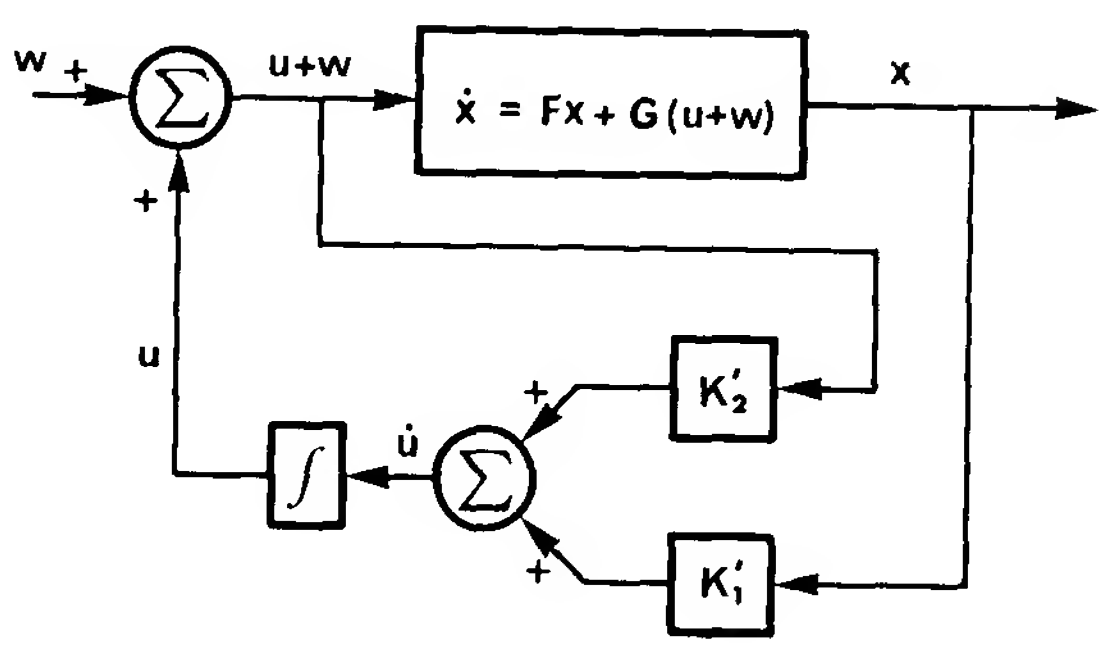
$$\begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix}$$

is completely observable for any D_{11} such that $D_{11}D'_{11} = Q$, and any D_{22} such that $D_{22}D'_{22} = R$, then the closed-loop system

$$\begin{bmatrix} \dot{x} \\ \dot{u}^* + \dot{w} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \dot{u}^* \end{bmatrix} = \begin{bmatrix} F & G \\ K'_1 & K'_2 \end{bmatrix} \begin{bmatrix} x \\ u^* + w \end{bmatrix} \qquad \begin{bmatrix} x(t_0) \\ u^*(t_0) + w \end{bmatrix} = \begin{bmatrix} 0 \\ w \end{bmatrix} \tag{10.3-10}$$

is asymptotically stable. Useful sufficient conditions are that $[F, D_{11}]$ be completely observable and R be positive definite.

The important point to observe from the asymptotically stable equation (10.3-10) is that the equilibrium condition is $x = 0$, $\dot{x} = 0$ and $u^* = -w$, as desired. Figure 10.3-1(a) shows the scheme for realizing the control law as expressed in (10.3-5), whereas Fig. 10.3-1(b) gives the scheme for the same



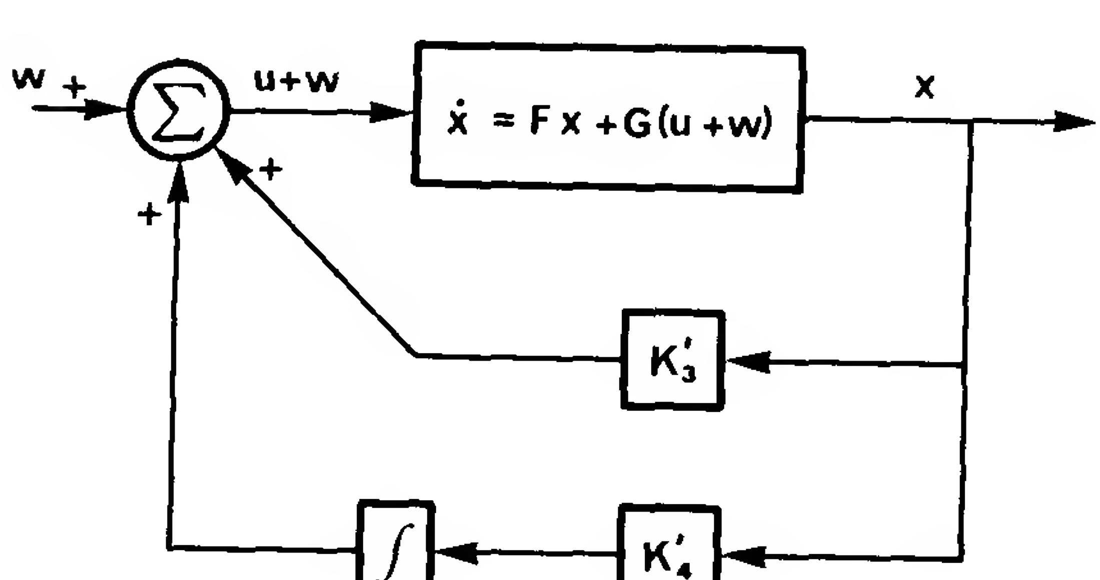


Fig. 10.3-1 Optimal regulator configurations.

control but expressed in the alternative form (10.3-6). Note that the second scheme is literally proportional-plus-integral (state) feedback, whereas the first scheme consists of integral feedback of the states and plant input. Clearly, if the input is not available, the first scheme cannot be used.

We stress again that through use of either of the forms of the optimal linear regulator with input derivative constraints shown in Fig. 10.3-1, the states of the original system (10.3-3) will approach an equilibrium value of zero asymptotically irrespective of the value of the constant input disturbance. The control will approach the negative value of the input disturbance asymptotically, thus cancelling its effect.

Now that a construction procedure for linear optimal regulators that include integral feedback has been given, and it has been established that constant input disturbances are cancelled by the controller output as time approaches infinity, we ask if such regulators have the same desirable properties as the standard optimal regulator. The answer is yes. Some of the properties can be demonstrated immediately, because the modified regulator is simply a standard regulator in disguise. Clearly, by replacing Q, S, and R with $Qe^{2\alpha t}$, $Se^{2\alpha t}$ and $Re^{2\alpha t}$, respectively, or, equivalently, by replacing

$$\begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix}$$

by

$$\begin{bmatrix} F + \alpha I & G \\ 0 & \alpha I \end{bmatrix}$$

in the derivations, an optimal system can be obtained with a degree of stability of at least α (see Problem 10.3-1). Again, it is clear that the sensitivity properties of the standard regulator carry over almost immediately into the modified regulator (see Problem 10.3-2). The question as to whether either of the modified linear optimal regulators can tolerate nonlinearities at the plant input will now be considered.

Let us suppose that a nonlinearity is interposed at the plant input, as in Fig. 10.3-2(a) or 10.3-2(b), with property that it changes the nominal plant input σ to $\phi(\sigma)$, where $\phi(\cdot)$ is a continuous function of σ, satisfying the inequality constraint $(\frac{1}{2} + \epsilon)\sigma'\sigma \leq \sigma'\phi(\sigma)$ for arbitrary positive ϵ—a constraint familiar from earlier discussions. [For the case of a scalar σ, this requires the graph of $\phi(\sigma)$ to lie between straight lines of slope $\frac{1}{2}$ and ∞ passing through the origin. See Fig. 10.3-3.]

To study the stability properties of the closed-loop plant and controller scheme with the newly introduced nonlinearity, we study first the scheme of Fig. 10.3-2(c), known from standard regulator theory to be asymptotically stable. Now, the system of Fig. 10.3-2(c) may be arranged as a linear subsystem of transfer function matrix

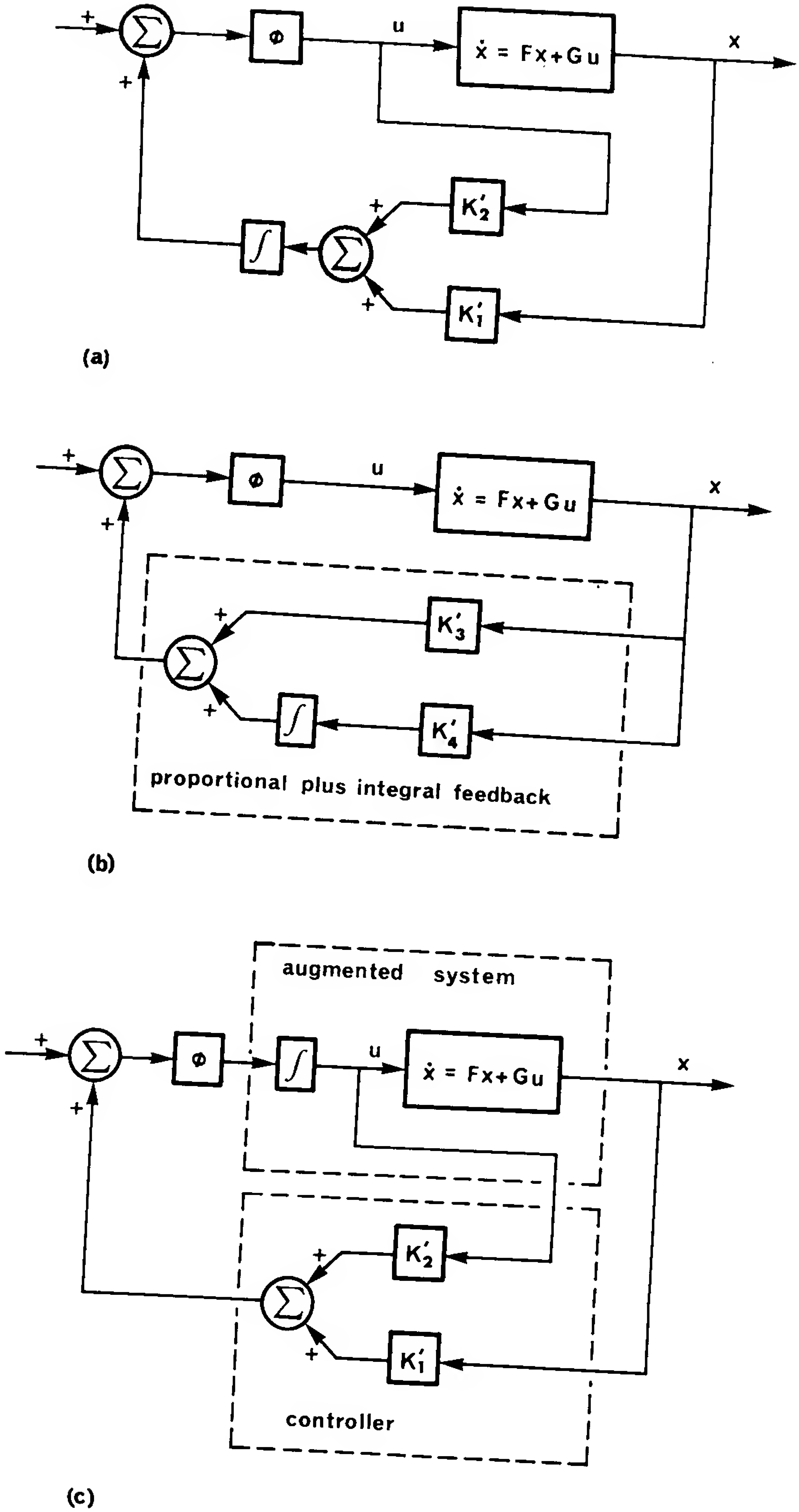


Fig. 10.3-2 Various forms for the one optimal regulator.

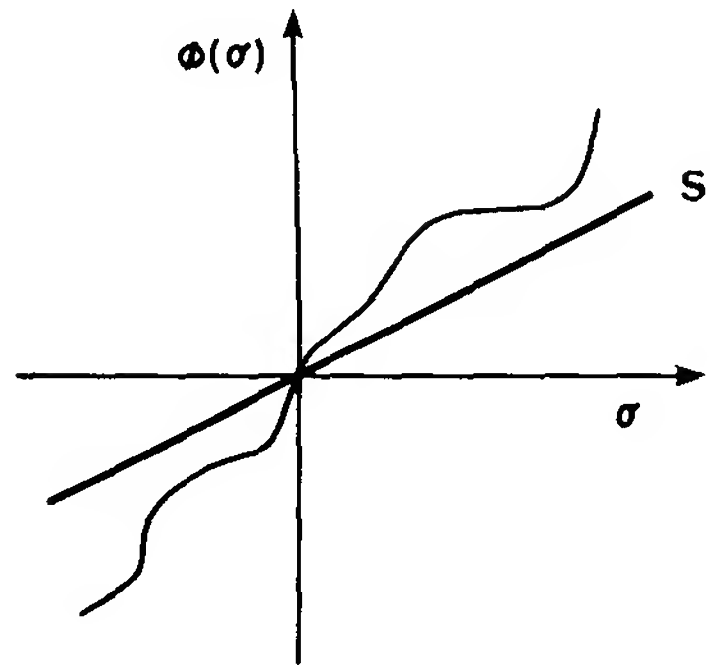


Fig. 10.3-3 Sector nonlinearities.

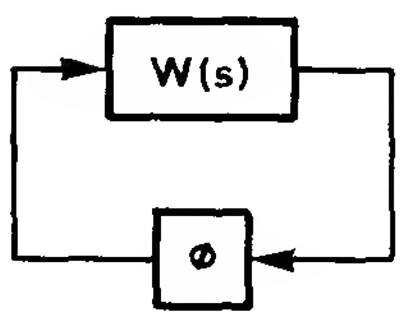


Fig. 10.3-4 Regulator with nonlinearities.

$$W(s) = s^{-1}K_1'(sI - F)^{-1}G + s^{-1}K_2' \qquad (10.3\text{-}11)$$

with feedback nonlinearities $\phi(\cdot)$, as indicated in Fig. 10.3-4. We claim now that the systems of Figs. 10.3-2(a) and 10.3-2(b) (which have the same form as the system of Fig. 10.3-4) are, in fact, identical to the system in Fig. 10.3-4, with the transfer function matrix of the linear part being $W(s)$, as in (10.3-11), in both cases. Consider first the arrangement of Fig. 10.3-2(a); the transfer function matrix of the linear subsystem is

$$K_1'(sI - F)^{-1}Gs^{-1} + K_2's^{-1} = W(s).$$

For the system of Fig. 10.3-2(b), the linear subsystem transfer function matrix is

$$\begin{aligned}(s^{-1}K_4' + K_3')(sI - F)^{-1}G &= (s^{-1}K_1' - s^{-1}K_3'F + K_3')(sI - F)^{-1}G \\ &= [s^{-1}K_1' + s^{-1}K_3'(sI - F)](sI - F)^{-1}G \\ &= s^{-1}K_3'G + s^{-1}K_1'(sI - F)^{-1}G \\ &= W(s).\end{aligned}$$

In the preceding development, we have used the definitions for K_3 and K_4 given in the previous section, Eq. (10.2-21).

We conclude that since all three systems of Fig. 10.3-2 can be arranged in the form of Fig. 10.3-4, with $W(s)$ given by (10.3-11), and since the system of Fig. 10.3-2(c) is known to be asymptotically stable, using standard regulator theory, then the optimal regulators of Figs. 10.3-2(a) and 10.3-2(b) are also asymptotically stable. Note, in particular, that for the linear optimal regulator

with proportional-plus-integral feedback, the nonlinearities are in the position of the plant input transducers. (A generalization of this result is requested in Problem 10.3-3.)

The preceding results also indicate that the gain margin and phase margin properties of single-input linear regulators, and their ability to tolerate time delays within the closed loop and still remain asymptotically stable, carry over to the modified regulators of Fig. 10.3-2.

To conclude the section, we discuss aspects of a second-order example. Consider the case when

$$F = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad G = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$S = 1$ and $R = 0$. To apply regulator theory to the system augmented by integrators at the inputs, we require that certain controllability and observability conditions be satisfied. Specifically, we require that $[F, G]$ be completely controllable and that the pair

$$\begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

be completely observable for any D such that $DD' = Q$. Since the matrices are time invariant, these conditions are perhaps most readily checked using the rank conditions, which are necessary and sufficient conditions for complete controllability and observability of time-invariant systems.

For $[F, G]$ to be completely controllable, the rank condition is

$$\text{Rank}\,(G \quad FG \quad F^2G \quad \cdots \quad F^{n-1}G) = n \tag{10.3-12}$$

where n is the order of the system. For $[F, D']$ to be completely observable, the rank condition is

$$\text{Rank}\,(D \quad F'D \quad (F')^2D \quad \cdots \quad (F')^{n-1}D) = n. \tag{10.3-13}$$

Applying these conditions to our example, we have

$$\text{Rank}\,(G \quad FG) = \text{Rank}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = 2$$

$$\begin{aligned}\text{Rank}\left(\begin{bmatrix} D' & 0 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} F' & 0 \\ G' & 0 \end{bmatrix}\begin{bmatrix} D' & 0 \\ 0 & 0 \end{bmatrix}\right) &= \text{Rank}\begin{bmatrix} D' & F'D' \\ 0 & G'D' \end{bmatrix} \\ &= \text{Rank}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= 3.\end{aligned}$$

We conclude that $[F, G]$ is completely controllable and the pair

$$\begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

is completely observable, as required.

To complete the solution of the problem, we may apply standard regulator theory in a straightforward manner to the augmented system, which is

$$\dot{x}_1 = F_1 x_1 + G_1 u_1$$

where $x_1 = [x' \quad u]'$ and $u_1 = \dot{u}$, with

$$F_1 = \begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad G_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The performance index is

$$V = \int_{t_0}^{\infty} (x_1' Q_1 x_1 + u_1' R_1 u_1)\, dt$$

where

$$Q_1 = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad R_1 = S = 1.$$

Hence, we find $\bar{P} = \lim_{t \to -\infty} P(t, T)$, where $P(\cdot, T)$ is the solution of

$$-\dot{P} = PF_1 + F_1'P - PG_1R_1^{-1}G_1'P + Q \qquad P(T, T) = 0.$$

The solution is

$$P = \left[\begin{array}{c:c} P_{11} & P_{21}' \\ \hdashline P_{21} & P_{22}' \end{array}\right] \left[\begin{array}{cc:c} 2 & 2 & 1 \\ 2 & 3 & 2 \\ \hdashline 1 & 2 & 2 \end{array}\right].$$

The optimal control to the augmented system is

$$u_1^* = -R_1^{-1}G_1'\bar{P}x_1 = -[1 \quad 2 \quad 2]x_1$$

or

$$\dot{u}^* = -[1 \quad 2]x - 2u,$$

but $\dot{x} = Fx + Gu$, and thus,

$$u = (G'G)^{-1}G'(\dot{x} - Fx) = [0 \quad 1]\dot{x}.$$

Combining the preceding two equations gives

$$\dot{u}^* = -[1 \quad 2]x - [0 \quad 2]\dot{x}.$$

The response of the system $\dot{x} = Fx + G(u + w)$ for a constant w with this control law (viz., $u^* = -[0 \quad 2]x - \int_0^t [1 \quad 2]x\, dt$) is plotted in Fig. 10.3-5(a).

The control u^* is plotted in Fig. 10.3-5(b). Corresponding quantities when a control law is derived from a standard regulator problem with

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$R = 1$ and $S = 0$, are given for comparison purposes in the same figures, using dotted lines. It is clear that the regulator with integral feedback can accommodate constant offset errors w, whereas the standard regulator cannot.

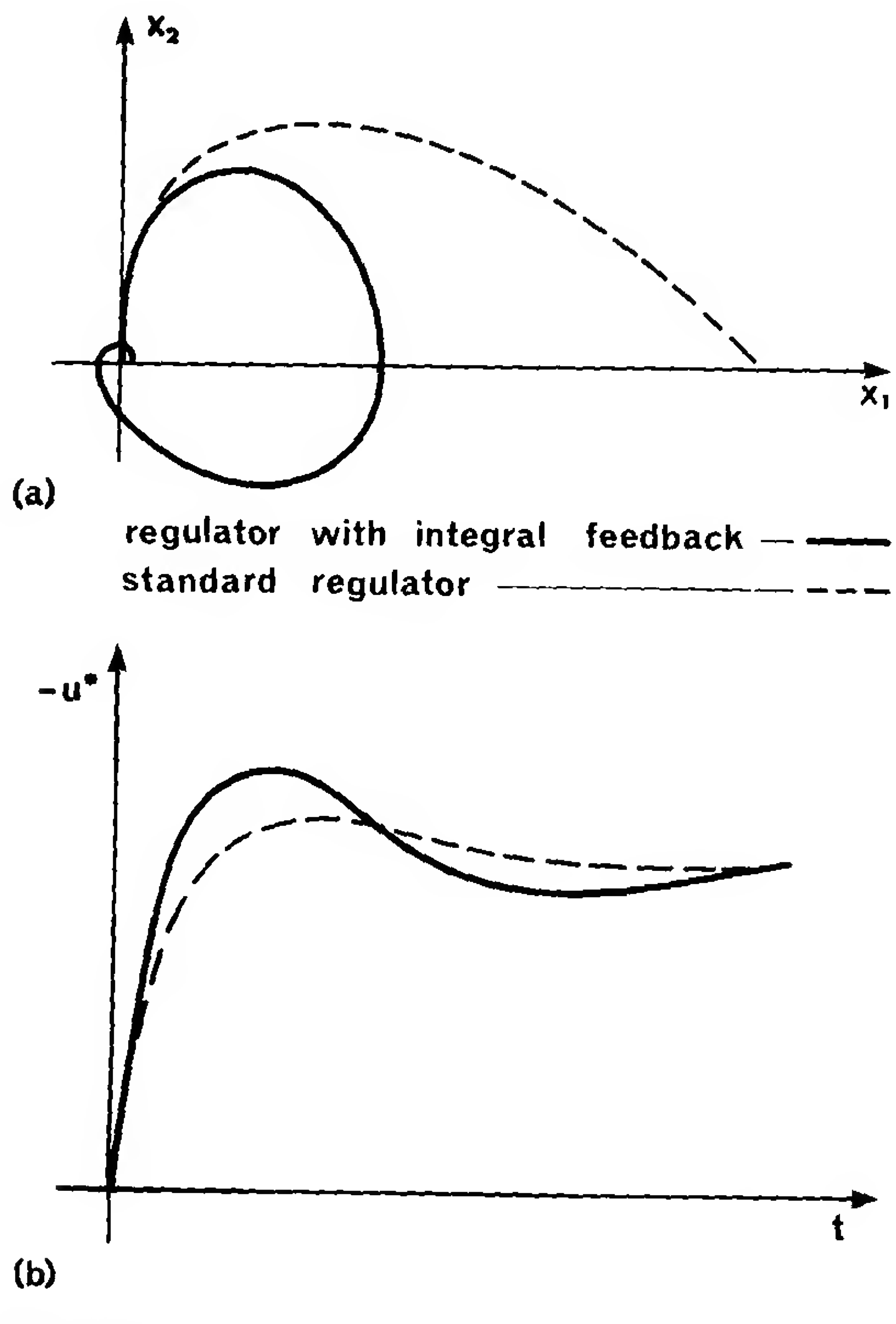


Fig. 10.3-5 Optimal control and response of regulators.

Problem 10.3-1. Repeat the derivation of the optimal linear regulator, which includes integral feedback for the case where the specifications require a guaranteed degree of stability of at least α.

Problem 10.3-2. Apply the sensitivity results of Chapter 7 to the optimal linear regulator which includes integral feedback, and thereby determine its sensitivity properties.

Problem 10.3-3. Consider the system $\dot{x} = Fx + Gu$, where the optimal control u satisfies the differential equation ($u^{(p)} = d^p u/dt^p$)

$$u^{(p)} = \sum_0^{p-1} K_i u^{(i)} + K_p x.$$

An ancillary property of optimality is that when $u^{(p)}$ is replaced by $\phi(u^{(p)})$, where ϕ represents nonlinearities as indicated in the text, the closed-loop system is asymptotically stable. Applying the relationship

$$u^{(i)} = (G'G)^{-1}G'(x^{(i+1)} - Fx^{(i)}),$$

we may rewrite the preceding control law as

$$u^{(p)} = \sum_0^{p-1} K_i(G'G)^{-1}G'x^{(i+1)} - \sum_0^{p-1} K_i(G'G)^{-1}G'Fx^{(i)} + K_p x.$$

Using this law, show that the closed-loop system is asymptotically stable, with u replaced by $\phi(u)$ [rather than $u^{(p)}$ replaced by $\phi(u^{(p)})$]. (*Hint:* This is a generalization of a result developed in the section. To solve the problem, equate transfer functions, as is done in the text for the case $p = 1$.)

REFERENCES

[1] Salukadze, M. E., "The Analytical Design of Optimal Controls in the Case of Constantly Acting Disturbances," *Automation and Remote Control*, Vol. 23, No. 6, December 1962, pp. 657–667.

[2] Kushner, H. J., "On the Status of Optimal Control and Stability for Stochastic Systems," *IEEE Intern. Conv. Record*, Vol. 14, Pt. 6, 1966, pp. 143–151.

[3] Johnson, C. D., "Optimal Control of the Linear Regulator with Constant Disturbances," *IEEE Trans. Auto. Control*, Vol. AC–13, No. 4, August 1968, pp. 416–421.

[4] Moore, J. B., and B. D. O. Anderson, "Optimal Linear Control Systems with Input Derivative Constraints," *Proc. IEE*, Vol. 114, No. 12, December 1967, pp. 1987–1990.

[5] Moore, J. B., "Tolerance of Nonlinearities in Linear Optimal Regulators with Integral Feedback," *Electron. Letters*, Vol. 5, No. 17, August 1969, pp. 409–410.

[6] Pearson, J. B., "Compensator Design for Dynamic Optimization," *Intern. J. Control*, Vol. 9, No. 4, 1969, pp. 473–482.

[7] Pearson, J. B., and C. Y. Ding, "Compensator Design for Multivariable Linear Systems," *IEEE Trans. Auto. Control*, Vol. AC–14, No. 2, April 1969, pp. 130–134.

[8] Ferguson, J. D., and Z. V. Rekasius, "Optimal Linear Control Systems with Incomplete State Measurements," *IEEE Trans. Auto. Control*, Vol. AC–14, No. 2, April 1969, pp. 135–140.

CHAPTER 11

TRACKING SYSTEMS

11.1 THE PROBLEM OF ACHIEVING A DESIRED TRAJECTORY

In previous chapters, the regulator problem—viz., the problem of returning a system to its zero state in some optimal fashion—is considered. This problem is, in fact, a special case of a wider class of problems where it is required that the outputs of a system follow or track a desired trajectory in some optimal sense. For the regulator, the desired trajectory is, of course, simply the zero state. In this chapter, we apply regulator theory and give extensions to solve the wider class of control problems that involves achieving a desired trajectory.

It is convenient to refer to these problems by one of three technical terms, the particular term used depending on the nature of the desired trajectory. If the plant outputs are to follow a *class* of desired trajectories, e.g., all polynomials up to a certain order, the problem is referred to as a *servo* (servomechanism) *problem*; if the desired trajectory is a *particular* prescribed function of time, the problem is called a *tracking problem*. When the outputs of the plant are to follow the response of another plant (or model) to either a specific command input or class of command inputs, the problem is referred to as the *model-following problem*.

The remainder of this section is devoted to a discussion of considerations common to all three of these problems, with particular attention being given to the selection of a performance index.

We recall that in selecting a performance index for a regulator, cost terms are constructed for the control energy and the energy associated with

the states. More specifically, for the linear system

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{11.1-1}$$

the principal performance index adopted throughout the book is the quadratic index

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} (u'Ru + x'Qx)\, dt \tag{11.1-2}$$

where R is some positive definite matrix and Q is some nonnegative definite matrix (the matrices being of appropriate dimensions). The quadratic nature of the cost terms ensures that the optimal control law is linear, and the constraints on the matrices Q and R ensure that the control law leads to a finite control.

When attempting to control the system (11.1-1) such that its output $y(\cdot)$ given by

$$y = H'x \tag{11.1-3}$$

tracks a desired trajectory $\tilde{y}(\cdot)$, there clearly should be a cost term in the performance index involving the error $(y - \tilde{y})$. A performance index that comes to mind immediately as a natural extension of the index (11.1-2) is the following:

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} [u'Ru + (y - \tilde{y})'Q(y - \tilde{y})]\, dt \tag{11.1-4}$$

where Q and R are nonnegative definite and positive definite matrices, respectively. Once again, we have quadratic terms that, as the next sections show, give rise to linear control laws.

Another objective, apart from minimizing the error $(y - \tilde{y})$ and the control u, may be to achieve a smooth system output y. It may be that the error $(y - \tilde{y})$ is not critical at every instant in time, but a smooth response is critical. One possible method that may be used to achieve the objective of smoothness is to generalize the index (11.1-4) as follows:

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} [u'Ru + x'Q_1x + (y - \tilde{y})'Q_2(y - \tilde{y})]\, dt \tag{11.1-5}$$

where Q_1 and Q_2 are nonnegative definite symmetric matrices.

The term $x'Q_1x$ tends to constrain the system states to be small, and thereby to encourage a smooth response. (A response that is not smooth must contain significant amounts of the output derivatives; thus, at least some of the system states must accordingly be large.) Unfortunately, without a restriction on Q_1, the two objectives represented by the second and third terms in (11.1-5) may be conflicting, for the following reason: The term $x'Q_1x$ is to encourage the entries of x to be small, and, in particular, the entries of $y = H'x$ to be small, whereas the term $(y - \tilde{y})'Q_2(y - \tilde{y})$ is to

encourage the entries of $y = H'x$ to be close to those of $\tilde{y}$, which, in general, will not be small.

We now claim that this difficulty can be overcome by restricting Q_1 to being of the form

$$Q_1 = [I - H(H'H)^{-1}H']'Q_3[I - H(H'H)^{-1}H'] \tag{11.1-6}$$

where Q_3 is an arbitrary nonnegative definite symmetric matrix. [It is implicitly assumed here that $(H'H)^{-1}$ exists. This restriction is equivalent to the very reasonable assumption arising from practical considerations, that the system outputs are linearly independent.]

To establish the preceding claim, we first write x as the sum of two orthogonal components x_1 and x_2, such that x_1 is in the range space of H'. That is, we write

$$x = x_1 + x_2 \tag{11.1-7}$$

where

$$Hl = x_1 \quad \text{and} \quad H'x_2 = 0 \tag{11.1-8}$$

for some vector l. These equations imply that

$$y = H'x = H'x_1, \tag{11.1-9}$$

which, in turn, implies that constraining y to follow $\tilde{y}$ constrains x_1, via (11.1-9), whereas no constraint on x_2 is implied. We conclude that any pair of objectives represented by cost terms $x_2'Q_3x_2$ and $(y - \tilde{y})'Q_2(y - \tilde{y})$ would not be in conflict for any Q_3, whereas, as we have already seen, a pair of objectives represented by cost terms $x'Qx$ and $(y - \tilde{y})'Q_2(y - \tilde{y})$ may be in conflict.

But now it is easy to check that the vector l as defined is

$$l = (H'H)^{-1}H'x \tag{11.1-10}$$

and, accordingly,

$$x_2 = x - Hl = [I - H(H'H)^{-1}H']x. \tag{11.1-11}$$

Therefore, Eq. (11.1-6) implies that $x'Q_1x = x_2'Q_3x_2$. Thus, since we have shown that the terms $x_2'Q_3x_2$ and $(y - \tilde{y})'Q_2(y - \tilde{y})$ are not conflicting in the sense implied by the previous discussion, and since $x_2'Q_3x_2 = x'Q_1x$ for the case when Q_1 is defined as in (11.1-6), then the performance index (11.1-5) with Q_1 given by (11.1-6) and Q_3 arbitrary has three cost terms that are not in direct conflict with one another. The first term is the control cost, the second is a "smoothness" cost, and the third is an error cost.

We now show that the index (11.1-5) with Q_1 defined by (11.1-6) may be written in an interesting form that will prove convenient to use in the next sections. More precisely, we shall show that the index (11.1-5) may be written

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} [u'Ru + (x - \tilde{x})'Q(x - \tilde{x})]\, dt \tag{11.1-12}$$

where

$$Q = Q_1 + HQ_2H' \tag{11.1-13}$$

$$\tilde{x} = H[H'H]^{-1}\tilde{y}. \tag{11.1-14}$$

The interpretation of the terms in the preceding index is straightforward enough. It appears that the cost terms of the index (11.1-5) involving the state and the error between the system output and the desired output are replaced by a single term involving the error between the state and some special state trajectory $\tilde{x}$, related to the desired output trajectory $\tilde{y}$. We have from (11.1-14) that $H'\tilde{x} = \tilde{y}$, and thus, if by some means $\tilde{x}$ were to become the state trajectory of (11.1-1), then the system output given by (11.1-3) would, in fact, be the desired output trajectory $\tilde{y}$. What characterizes the trajectory $\tilde{x}$ is the important property, to be proved, that its component $\tilde{x}_2$ in the null space of H' is zero. In other words, if the system output vector y were augmented by another vector output $\bar{y} = \bar{H}x$, with each entry of $\bar{y}$ linearly independent of the entries of y *and* with the combined dimension of y and $\bar{y}$ equal to the state-vector dimension, then with $\tilde{x}$ as the state trajectory this output $\bar{y}$ would be zero. The specified properties of $\tilde{x}$ suggest that it be referred to as the *desired state trajectory*.

We have yet to prove that $\tilde{x}_2$, the component of $\tilde{x}$ in the null space of H', is zero, and also that the indices (11.1-5) and (11.1-12) are, in fact, identical. Using reasoning similar to that used in deriving (11.1-11), we have that

$$\begin{aligned}\tilde{x}_2 &= [I - H(H'H)^{-1}H']\tilde{x} \\ &= \tilde{x} - H(H'H)^{-1}\tilde{y} \\ &= 0.\end{aligned}$$

This result yields immediately that $\tilde{x}'Q_1\tilde{x} = 0$ and, in fact, $Q_1\tilde{x} = 0$, where Q_1 is defined as in (11.1-6). It then follows that

$$\begin{aligned}(x - \tilde{x})'Q(x - \tilde{x}) &= x'Q_1x + (x - \tilde{x})'HQ_2H'(x - \tilde{x}) \\ &= x'Q_1x + (y - \tilde{y})'Q_2(y - \tilde{y})\end{aligned}$$

where Q is defined as in (11.1-13). Our claim is established.

For any particular application of the performance index just developed, selections have to be made of the matrices Q_2, Q_3, and R. Also, a selection may have to be made of the terminal time T. It may be necessary to try a range of values of these quantities and to select the particular one which is most appropriate for any given situation. A case of particular interest is the limiting case as the terminal time T approaches infinity. For this case, when all the matrices are time invariant, part, if not all, of the optimal controller becomes time invariant. However, difficulties may arise for this case because it may not be possible for the plant to track the desired trajectories so that the error approaches zero as time becomes infinite with a finite control law.

Moreover, even if it is possible to track in this sense using a finite control law, unless the control approaches zero as time approaches infinity, other difficulties arise due to the fact that the performance index would be infinite for all controls and therefore attempts at its minimization would be meaningless.

The next section considers finite terminal-time servo-, tracking-, and model-following problems. The final section considers the limiting situation as the integration interval becomes infinite. Of particular interest in this section are the cases where the desired trajectory is either a step, ramp, or parabolic function. Key references for the material that follows are [1] through [3].

11.2 FINITE-TIME RESULTS

The servo problem. As stated in the previous section, the servo problem is the task of controlling a system so that the system output follows a prescribed signal, where all that is known about the signal is that it belongs to a known class of signals. We consider a particular specialization.

Optimal servo problem. Suppose we are given the n-dimensional linear system having state equations

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{11.2-1}$$

$$y = H'x \tag{11.2-2}$$

where the m entries of y are linearly independent or, equivalently, the matrix H has rank m. Suppose we are also given an m-vector incoming signal $\tilde{y}$, which is the output of the known p-dimensional linear system

$$\dot{z} = Az \tag{11.2-3}$$

$$\tilde{y} = C'z \tag{11.2-4}$$

for some initial state $z(t_0)$. Without loss of generality, the pair $[A, C]$ is completely observable. The optimal servo problem is to find the optimal control u^* for the system (11.2-1), such that the output y tracks the incoming signal $\tilde{y}$, minimizing the index

$$\begin{aligned} V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} \{u'Ru & \\ & + x'[I - H(H'H)^{-1}H']'Q_1[I - H(H'H)^{-1}H']x \\ & + (y - \tilde{y})'Q_2(y - \tilde{y})\}\, dt \end{aligned} \tag{11.2-5}$$

where Q_1 and Q_2 are nonnegative definite symmetric matrices, and R is positive definite symmetric. (As usual, all the various matrices are assumed to have continuous entries.)

Observe that we are requiring that our desired trajectory $\tilde{y}$ be a solution to a linear differential equation. This, of course, rules out trajectories $\tilde{y}$ which have discontinuities for $t > t_0$. We also note that the special case when C is a vector and A is given by

$$A = \begin{bmatrix} 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & 0 & 1 & & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot \\ \cdot & & & & \cdot & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}$$

leads to the class of $\tilde{y}$ consisting of all polynomials of degree $(p - 1)$.

Proceeding with the solution of the servo problem, we note an equivalent expression for the index (11.2-5), derived in the previous section—namely,

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} [u'Ru + (x - \tilde{x})'Q(x - \tilde{x})]\, dt \qquad (11.2\text{-}6)$$

where

$$Q = [I - H(H'H)^{-1}H']'Q_1[I - H(H'H)^{-1}H'] + HQ_2H' \qquad (11.2\text{-}7)$$

and $\tilde{x}$, the desired state trajectory, is

$$\tilde{x} = H(H'H)^{-1}\tilde{y}. \qquad (11.2\text{-}8)$$

Throughout the book, various minimization problems are solved by first applying a transformation to convert the minimization problem to a standard regulator problem. The standard regulator results are then interpreted, using the transformations to give a solution to the original minimization problem. This will also be our method of attack here. To convert the preceding servo problem to a regulator problem, we require the following assumption.

ASSUMPTION 11.2-1. The state z is directly measurable. (Later, we shall discuss the elimination of this assumption.)

We now define a new variable

$$\hat{x} = \begin{bmatrix} x \\ z \end{bmatrix} \qquad (11.2\text{-}9)$$

and new matrices

$$\hat{F} = \begin{bmatrix} F & 0 \\ 0 & A \end{bmatrix} \quad \hat{G} = \begin{bmatrix} G \\ 0 \end{bmatrix}$$

$$\hat{Q} = \begin{bmatrix} Q & -QH(H'H)^{-1}C' \\ -C(H'H)^{-1}H'Q & C(H'H)^{-1}H'QH(H'H)^{-1}C' \end{bmatrix}. \qquad (11.2\text{-}10)$$

These variables and matrices are so constructed that when applied to the problem of minimizing (11.2-6) with the relationships (11.2-1) and (11.2-3) holding, we have the standard regulator problem requiring minimization of the quadratic index

$$V(\hat{x}(t_0), u(\cdot), T) = \int_{t_0}^{T} (u'Ru + \hat{x}'\hat{Q}\hat{x})\, dt \tag{11.2-11}$$

with the relationship

$$\dot{\hat{x}} = \hat{F}\hat{x} + \hat{G}u \qquad \hat{x}(t_0) \quad \text{given} \tag{11.2-12}$$

holding. This result may be checked by the reader as an exercise.

Applying the regulator theory of Chapter 3, Sec. 3.3, to the minimization problem, (11.2-11) and (11.2-12), gives immediately that the optimal control u^* is

$$u^* = -R^{-1}\hat{G}'\hat{P}\hat{x} \tag{11.2-13}$$

where $\hat{P}(\cdot)$ is the solution of the Riccati equation

$$-\dot{\hat{P}} = \hat{P}\hat{F} + \hat{F}'\hat{P} - \hat{P}\hat{G}R^{-1}\hat{G}'\hat{P} + \hat{Q} \qquad \hat{P}(T) = 0. \tag{11.2-14}$$

The minimum index is

$$V^*(\hat{x}(t_0), T) = \hat{x}'(t_0)\hat{P}(t_0)\hat{x}(t_0). \tag{11.2-15}$$

We now interpret these results in terms of the variables and matrices of the original problem, using the definitions (11.2-9) and (11.2-10). First, we partition $\hat{P}$ as

$$\hat{P} = \begin{bmatrix} P & P'_{21} \\ P_{21} & P_{22} \end{bmatrix} \tag{11.2-16}$$

where P is an $n \times n$ matrix. Substituting (11.2-10) and (11.2-16) into (11.2-13) gives the optimal control u^* as

$$u^* = K'x + K'_1 z \tag{11.2-17}$$

where

$$K' = -R^{-1}G'P \tag{11.2-18}$$

$$K'_1 = -R^{-1}G'P'_{21}. \tag{11.2-19}$$

The Riccati equation (11.2-14) becomes now

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \tag{11.2-20}$$

$$-\dot{P}_{21} = P_{21}F + A'P_{21} - P_{21}GR^{-1}G'P - C(H'H)^{-1}H'Q \tag{11.2-21}$$

$$\begin{aligned} -\dot{P}_{22} = P_{22}A &+ A'P_{22} - P_{21}GR^{-1}G'P'_{21} \\ &+ C(H'H)^{-1}H'QH(H'H)^{-1}C' \end{aligned} \tag{11.2-22}$$

with the boundary conditions $P(T) = 0$, $P_{21}(T) = 0$, and $P_{22}(T) = 0$. The minimum index is

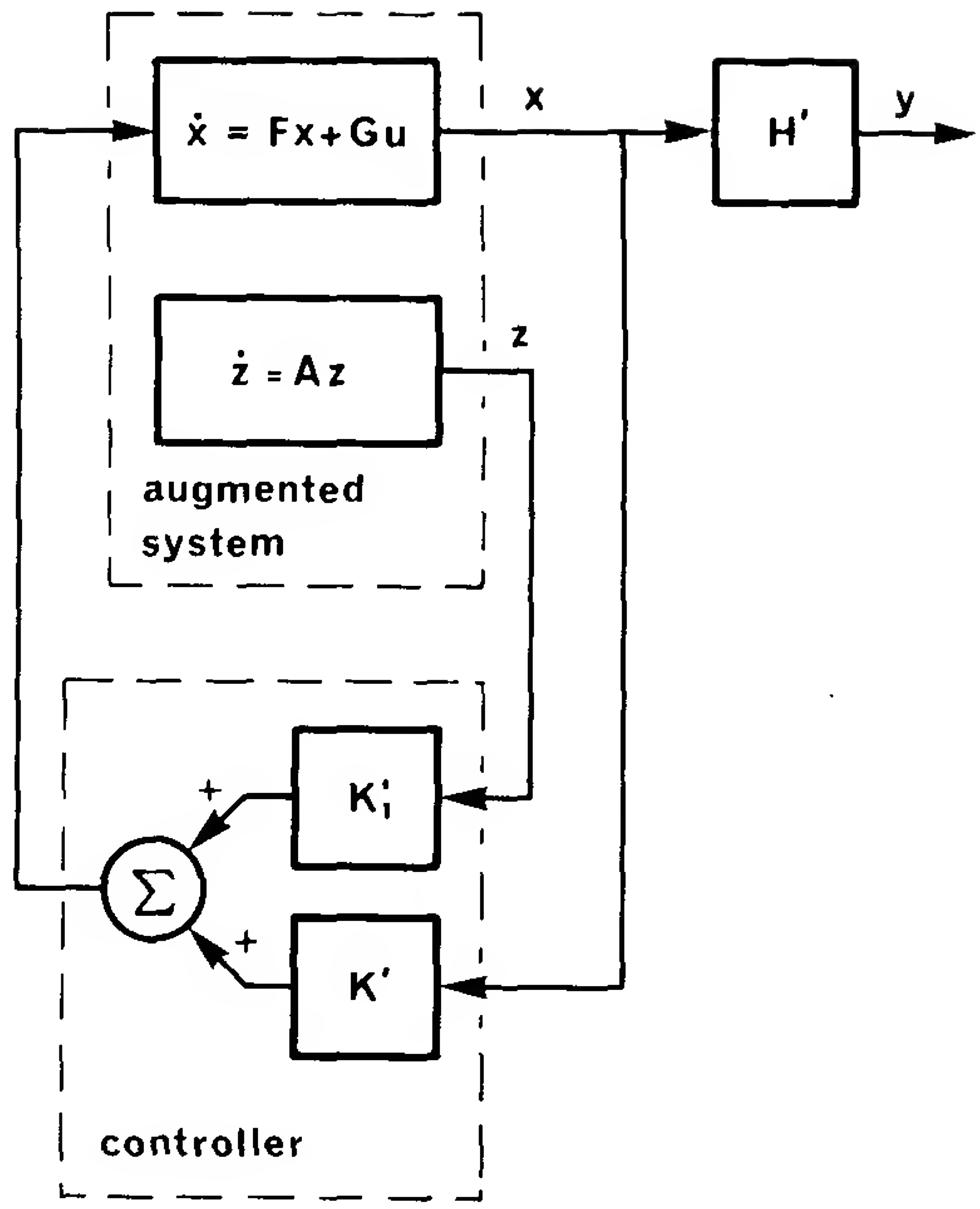


Fig. 11.2-1 Regulator control of augmented system.

$$V^*(x(t_0), T) = x'(t_0)P(t_0)x(t_0) + 2x'(t_0)P'_{21}(t_0)z(t_0)$$
$$+ z'(t_0)P_{22}(t_0)z(t_0). \tag{11.2-23}$$

Figure 11.2-1 shows the augmented system (11.2-12) separated into its component systems (11.2-1) and (11.2-3) and controlled by use of linear state-variable feedback, as for a standard regulator. Figure 11.2-2 shows the same system redrawn as the solution to the servo problem. We observe that it has the form of a regulator designed by minimizing the index

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} (u'Ru + x'Qx)\, dt$$

for the system (11.2-1) using regulator theory. In addition, there is an external input, which is the state z of the linear system (11.2-3) and (11.2-4). *The feedback part of the control is independent of A, C, and $z(t_0)$.*

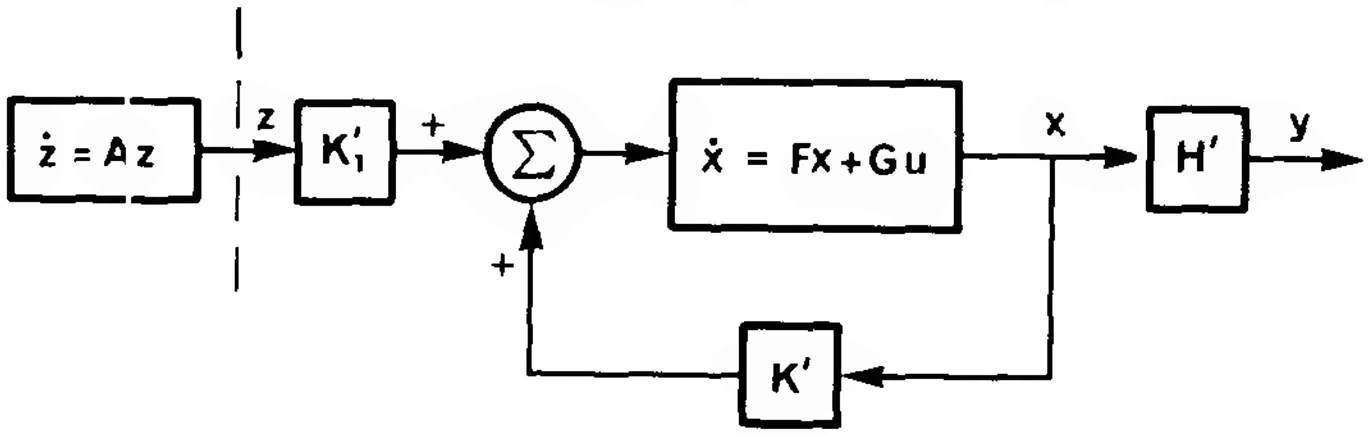


Fig. 11.2-2 Rearrangement of regulator system of Fig. 11.2-1.

The results to this point depend on Assumption 11.2-1, namely, that the state z is directly measurable. Certainly, if z is available, the servo problem is solved. However, often in practice, only an incoming signal $\tilde{y}$ is at hand. For this case, a state estimator may be constructed with $\tilde{y}$ as input and, of course, an estimate $\hat{z}$ of z as output, since the pair $[A, C]$ appearing in (11.2-3) and (11.2-4) is completely observable. The estimator may be constructed by using the results of Chapters 8 and 9, with $\hat{z}$ approaching z arbitrarily fast, at least if A and C are constant. The resulting system is shown in Fig. 11.2-3(a).

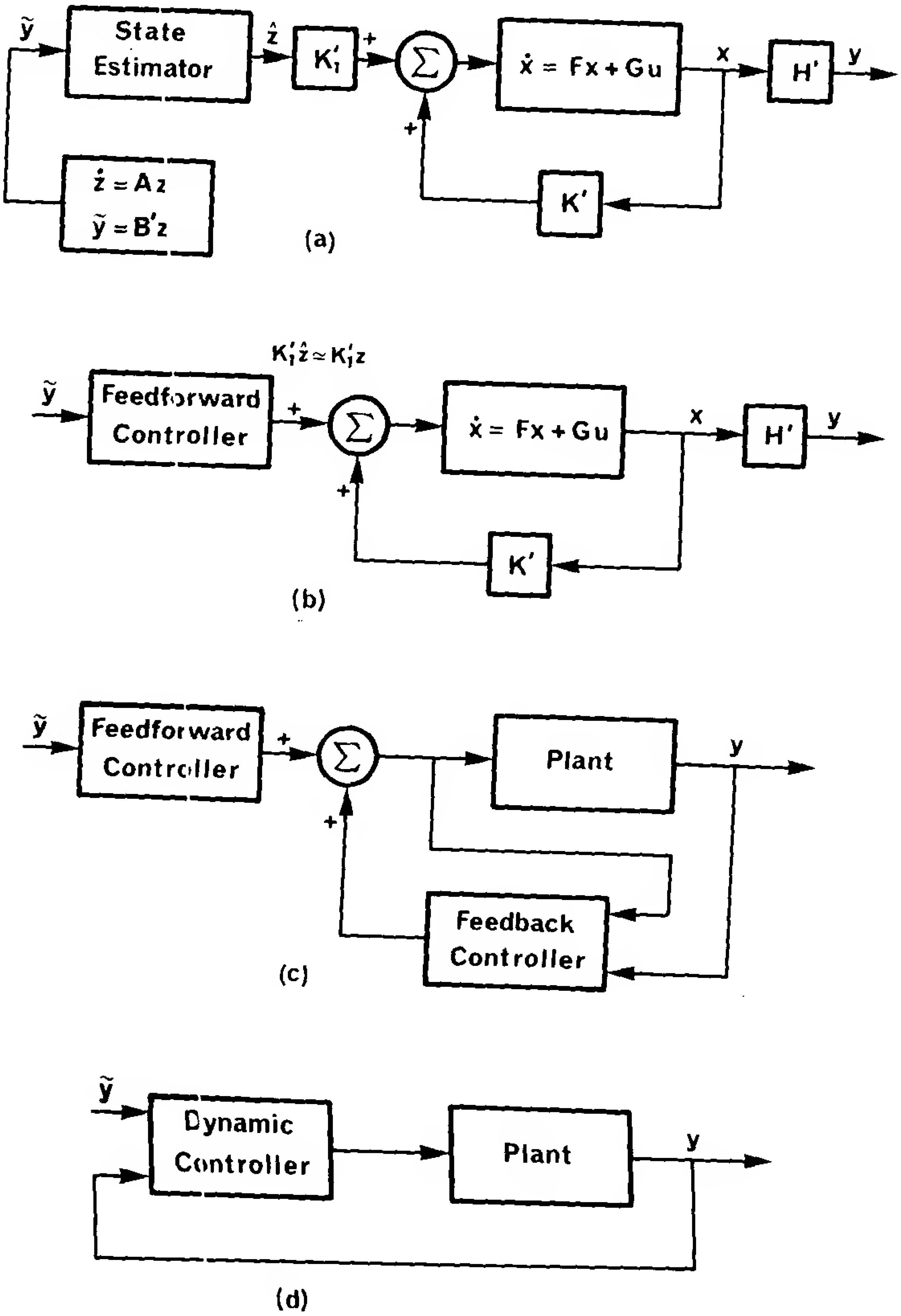


Fig. 11.2-3 Four forms for the optimal servo-system.

It is redrawn in Fig. 11.2-3(b) to illustrate that the state estimator and control law K_1' can be combined to give a (dynamic) feedforward controller.

Likewise, if the state x of the plant is not directly measurable, the memoryless linear state feedback may be replaced by a dynamic controller, which estimates the state x and then forms the appropriate linear transformation of this estimate, always assuming complete observability of $[F, H]$. Figure 11.2-3(c) shows this arrangement. Figure 11.2-3(d) shows one further possibility, where the estimate of x and z is carried out simultaneously in the one estimator; this arrangement may yield a reduction in the dimension of the linear system comprising the controller. We shall now summarize the optimal servo problem solution.

Solution to the finite-time optimal servo problem. For the systems (11.2-1), (11.2-2), and (11.2-3), (11.2-4), and performance index (11.2-5), the optimal control u^* is given by (11.2-17) through (11.2-22). The minimum index is given in (11.2-23). With Assumption 11.2-1 holding, the form of the optimal controller is indicated in Fig. 11.2-2(a). If only an estimate $\hat{z}$ of the state z is available, then an approximation to the optimal controller is as indicated in Fig. 11.2-3(b). The closer the estimate $\hat{z}$ is to z, the better is the approximation. Further possibilities are shown in Figs. 11.2-3(c) and 11.2-3(d), where estimation of x is required.

The tracking problem. It may be that the desired trajectory $\tilde{y}(t)$ for all t in the range $t_0 \leq t \leq T$ is known a priori. If this is the case, the contruction of a state estimator to determine an estimate $\hat{z}$ of z is unnecessary. This represents a considerable saving if $\tilde{y}$ is, e.g., the output of a high-order, time-varying, linear system. We now define the optimal tracking problem and give its solution.

Optimal tracking problem. Suppose we are given the n-dimensional linear system having state equations

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{11.2-1}$$

$$y = H'x \tag{11.2-2}$$

where the m entries of y are linearly independent. Suppose we are also given an m vector $\tilde{y}(t)$ for all t in the range $t_0 \leq t \leq T$ for some times t_0 and T with $t_0 < T$. The optimal tracking problem is to find the optimal control u^* for the system (11.2-1), such that the output y tracks the signal $\tilde{y}$, minimizing the index

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} \{u'Ru + x'[I - H(H'H)^{-1}H']'Q_1[I - H(H'H)^{-1}H']x + (y - \tilde{y})'Q_2(y - \tilde{y})\}\, dt \tag{11.2-5}$$

where Q_1 and Q_2 are nonnegative definite symmetric and R is positive definite symmetric.

We first make the following temporary assumption.

TEMPORARY ASSUMPTION 11.2-2. The vector $\tilde{y}(t)$ for all t in the range $t_0 \le t \le T$ is the output of a linear finite dimensional system

$$\dot{z} = Az \tag{11.2-3}$$

$$\tilde{y} = C'z \tag{11.2-4}$$

with the pair $[A, C]$ *not* necessarily assumed to be completely observable.

With this assumption holding, the optimal control u^* is given using the optimal servo results (11.2-17) through (11.2-22) as

$$u^* = K'x + u_{\text{ext}} \tag{11.2-24}$$

where

$$u_{\text{ext}} = -R^{-1}G'b. \tag{11.2-25}$$

The matrix K is calculated as before from (11.2-18) and (11.2-20), and the vector b is the product $(P'_{21}z)$. Moreover, the minimum index V^*, given from (11.2-23), may be written

$$V^*(x(t_0), T) = x'(t_0)P(t_0)x(t_0) + 2x'(t_0)b(t_0) + c(t_0) \tag{11.2-26}$$

where c replaces the term $z'P_{22}z$.

The value of the matrices P'_{21} and P_{22} and the vector z cannot be determined independently unless the matrices A and C are known. However, the products $b = P'_{21}z$ and $c = z'P_{22}z$ can be determined directly from $\tilde{y}(\cdot)$, as follows.

Differentiating the product $(P'_{21}z)$ and applying Eq. (11.2-21) for P'_{21} and (11.2-3) for z, we get

$$\begin{aligned}
-\frac{d}{dt}(P'_{21}z) &= -\dot{P}'_{21}z - P'_{21}\dot{z} \\
&= F'P'_{21}z + P'_{21}Az - PGR^{-1}G'P'_{21}z - QH(H'H)^{-1}C'z \\
&\quad - P'_{21}Az \\
&= (F - GR^{-1}G'P)'(P'_{21}z) - QH(H'H)^{-1}\tilde{y}
\end{aligned}$$

with the boundary condition $P'_{21}(T)z(T) = 0$, following from $P_{21}(T) = 0$. This means that with $\tilde{y}(t)$ known for all t in the range $t_0 \le t \le T$, the term b can be calculated from the linear differential equation

$$-\dot{b} = (F + GK')'b - Q\tilde{x} \qquad b(T) = 0 \tag{11.2-27}$$

where, of course,

$$\tilde{x} = H(H'H)^{-1}\tilde{y}. \tag{11.2-8}$$

The optimal control law (11.2-24) and (11.2-25) is therefore realizable without recourse to using z, or an estimate $\hat{z}$ of z: Eq. (11.2-27) is solved, backward in time, to determine $b(t)$, which is then used in the optimal control law implementation. Matrices A and C do not play any role in determining u^*.

An equation for $c(\cdot)$ is determined by differentiating $c = z'P_{22}z$ as follows:

$$\begin{aligned}\frac{d}{dt}(z'P_{22}z) &= z'\dot{P}_{22}z + 2z'P_{22}\dot{z}\\ &= z'P_{21}GR^{-1}G'P'_{21}z - z'C(H'H)^{-1}H'QH(H'H)^{-1}C'z.\end{aligned}$$

Using the equations (11.2-4) and (11.2-8) and the identifications $b = P'_{21}z$ and $c = z'P_{22}z$, we have that $c(\cdot)$ is the solution of the differential equation

$$\dot{c} = b'GR^{-1}G'b - \tilde{x}Q\tilde{x} \quad c(T) = 0. \tag{11.2-28}$$

We observe by using (11.2-27) and (11.2-28), that the matrices A and C do not play any role in determining V^* from (11.2-26).

Since the differential equations for $b(\cdot)$ and $c(\cdot)$ can, in fact, be solved without Temporary Assumption 11.2-2, we indicate in outline a procedure for verifying that the control given by (11.2-24) and (11.2-25) is, in fact, optimal without Assumption 11.2-2. With u^* defined by (11.2-24) and (11.2-25), but not assumed optimal, and with $V^*[x(t_0), T]$ defined as in (11.2-26)—again, of course, not assumed optimal—Problem 11.2-1 asks for the establishing of the following identity:

$$\begin{aligned}V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} (u - u^*)'R(u - u^*)\, dt\\ + V^*(x(t_0), T).\end{aligned} \tag{11.2-29}$$

Here, $u(\cdot)$ is an arbitrary control. Optimality of u^* and V^* is then immediate. The preceding results are now summarized.

Solution to the finite-time tracking problem. For the system (11.2-1) and (11.2-2), and performance index (11.2-5), with the desired trajectory $\tilde{y}(t)$ available for all t in the range $t_0 \leq t \leq T$, the optimal control u^* is given from

$$u^* = -R^{-1}G'(Px + b)$$

[see (11.2-24), (11.2-25), and (11.2-18)], where $P(\cdot)$ is the solution of

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \quad P(T) = 0$$

and $b(\cdot)$ is the solution of

$$-\dot{b} = (F - GR^{-1}G'P)'b + QH(H'H)^{-1}\tilde{y} \quad b(T) = 0$$

[see (11.2-27) and (11.2-8)]. The minimum index is

$$V^*(x(t_0), T) = x'(t_0)P(t_0)x(t_0) + 2x'(t_0)b(t_0) + c(t_0) \tag{11.2-26}$$

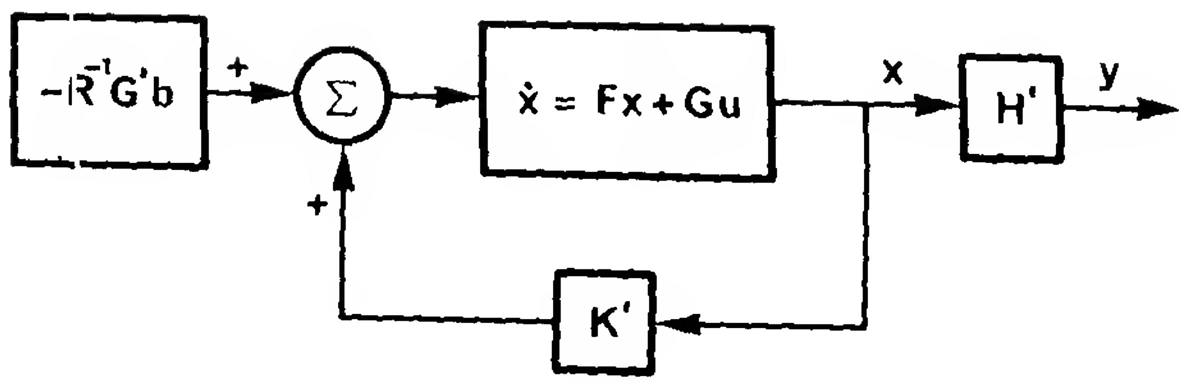


Fig. 11.2-4 Optimal tracking system.

where $c(t_0)$ is determined from

$$\dot{c} = b'GR^{-1}G'b - \tilde{y}'(H'H)^{-1}H'QH(H'H)^{-1}\tilde{y} \quad c(T) = 0.$$

The optimal controller is as indicated in Fig. 11.2-4.

The model-following problem. The model-following problem we shall consider is one that is easily reduced to a servo problem.

Optimal model-following problem. Find a control u^* for the linear system

$$\dot{x} = Fx + Gu \qquad x(t_0) \text{ given} \tag{11.2-1}$$

$$y = H'x \tag{11.2-2}$$

which minimizes the index

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} \{u'Ru + x'[I - H(H'H)^{-1}H']'Q_1[I - H(H'H)^{-1}H']x + (y - \tilde{y})'Q_2(y - \tilde{y})\}\, dt. \tag{11.2-5}$$

In (11.2-5), Q_1 and Q_2 are nonnegative definite symmetric, R is positive definite symmetric, and $\tilde{y}$ is the response of a linear system or model

$$\dot{z}_1 = A_1 z_1 + B_1 r \qquad z_1(t_0) \text{ given} \tag{11.2-30}$$

$$\tilde{y} = C_1' z_1 \tag{11.2-31}$$

to command inputs r, which, in turn, belong to the class of zero input responses of the system

$$\dot{z}_2 = A_2 z_2 \qquad z_2(t_0) \text{ given} \tag{11.2-32}$$

$$r = C_2' z_2 \tag{11.2-33}$$

as indicated in Fig. 11.2-5.

The case when the command signal r is known a priori may be solved by direct application of the tracking problem results; it is left to the reader (see Problem 11.2-4).

The two systems, (11.2-30) and (11.2-31), and (11.2-32) and (11.2-33),

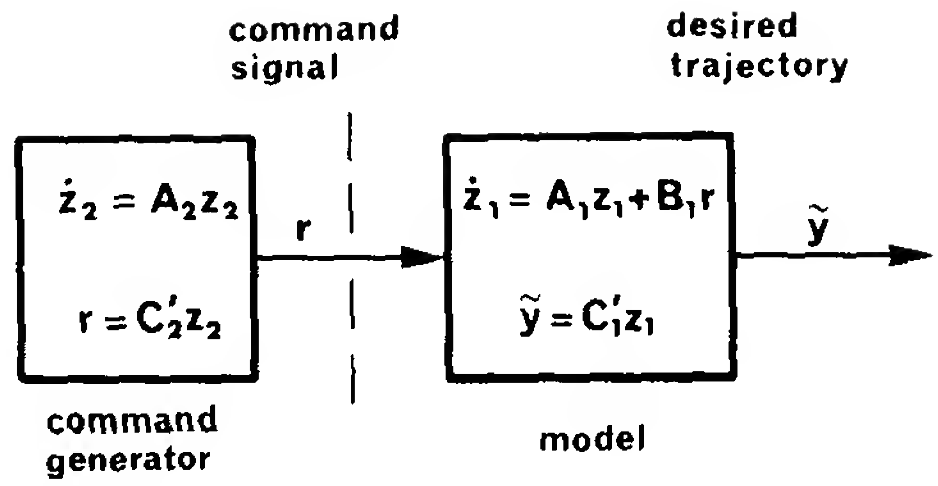


Fig. 11.2-5 Desired trajectory for model-following problem.

together form a linear system of the form of

$$\dot{z} = Az \tag{11.2-3}$$

$$\tilde{y} = C'z \tag{11.2-4}$$

where $z = [z_1' \quad z_2']'$ and the matrices A and C' are given from

$$A = \begin{bmatrix} A_1 & B_1 C_2' \\ 0 & A_2 \end{bmatrix} \qquad C' = [C_1' \quad 0]. \tag{11.2-34}$$

For the case when z_1 and z_2 are available, the equations for the solution to the model-following problem are identical to those for the servo problem, with A and C given by (11.2-34). In case of nonavailability of z_2, state estimation is required, again in the same manner as in the servo problem. Figure 11.2-6

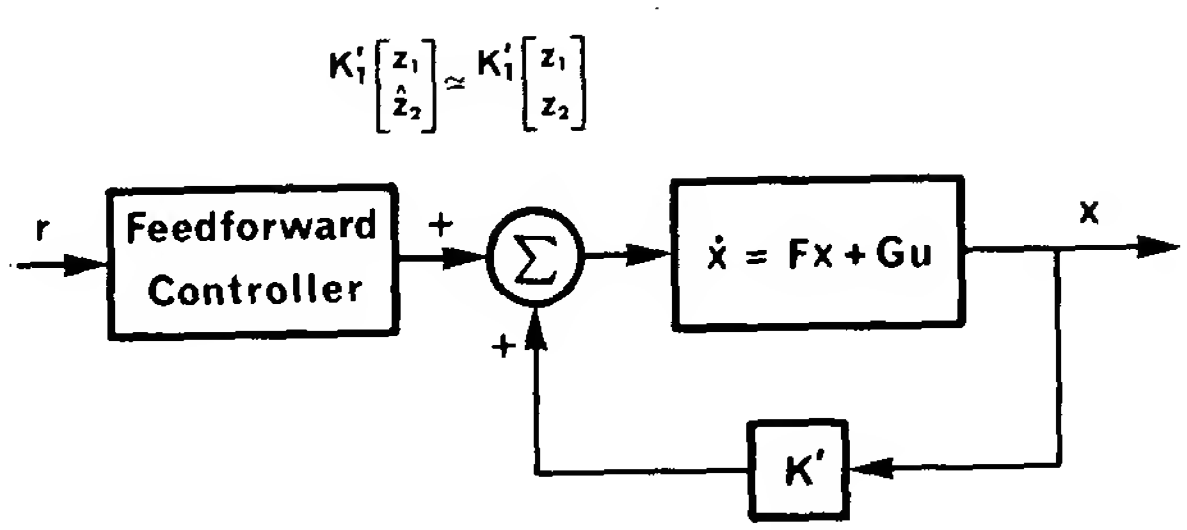


Fig. 11.2-6 Model-following system.

shows a model-following system with a feedforward controller consisting of a model to give z_1, and an estimator to give $\hat{z}_2 = z_2$.

Problem 11.2-1. Show that the index (11.2-5), or (11.2-6), may be written in the form of (11.2-29) where u^* is given from (11.2-24) and (11.2-25) and V^* is given from (11.2-26). What happens if the term $[x'(T) - \tilde{x}'(T)]D[x(T) - \tilde{x}(T)]$ is added to the performance index as a final-time cost term? Give the solution to the tracking problem for this case and verify your solution.

Problem 11.2-2. Derive the tracking problem results directly, using the Hamilton–Jacobi theory of Chapter 2 (i.e., without using the regulator theory results).

Problem 11.2-3. Extending the ideas of Chapter 10, give the form of the optimal servo and tracking systems if a term $\dot{u}'Z\dot{u}$ is included in the integrand of the performance index (11.2-5).

Problem 11.2-4. Give two forms of solution to the model-following problems as stated in the section, for the case when the command signal r is known a priori and construction of a state estimator for the command signal generator is out of the question.

11.3 INFINITE-TIME RESULTS

The finite-time results of the previous section yield servo systems, tracking systems, and model-following systems, which consist of a standard optimal regulator with an appropriate external input. We expect that by extending these results to the infinite-time case, the various systems will consist of an infinite-time standard optimal regulator, again with an appropriate external input. We recall that the optimal regulator for the infinite-time case has some very desirable engineering properties, such as the ability to tolerate certain classes of nonlinearities in the input transducers. For the case when the plant is time invariant and the parameters of the performance index are time invariant, the optimal regulator has a controller that is also time invariant. This is, of course, another desirable property from the engineering point of view. Therefore, the infinite-time servo, tracker, or model follower can be expected to possess a number of desirable properties, such as those just mentioned.

In this section, we restrict attention to plants and performance indices with time-invariant parameters, and we extend the finite-time results of the previous section to the infinite-time case. The aim is to achieve systems that track an input signal, and that also have the property that when the incoming signal is removed, the system has the same desirable properties as the standard optimal regulator. In particular, we require that the optimal system have time-invariant controllers.

We shall first consider the case of the completely controllable plant

$$\dot{x} = Fx + Gu \tag{11.3-1}$$

$$y = H'x \tag{11.3-2}$$

and the performance index

$$V(x(t_0), u(\cdot)) = \lim_{T\to\infty} \int_{t_0}^{T} \{u'Ru + (y - \bar{y})'Q(y - \bar{y})\}\, dt \tag{11.3-3}$$

where the matrices F, G, H, R, and Q are constant. The vector $\bar{y}$ is, as previously, the desired output trajectory. For the sake of simplicity, we have

omitted the "smoothing" term in the index. We also assume that the pair $[F, D]$ is completely observable for any D satisfying $DD' = HQH'$. We now discuss an important special case.

Infinite-time cases completely reducible to regulator problems. Consider first the case when the desired output trajectory $\tilde{y}$ is derived from an asymptotically stable system, and is given by

$$\dot{z} = Az, \qquad z(t_0) \text{ given} \tag{11.3-4}$$

$$\tilde{y} = C'z. \tag{11.3-5}$$

(We shall later remove the assumption of asymptotic stability.) In the previous section, we observed that the finite-time servo problem could be reformulated as a regulator problem for an augmented system

$$\begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} F & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} G \\ 0 \end{bmatrix} u. \tag{11.3-6}$$

Since $[F, G]$ is completely controllable, and since $\dot{z} = Az$ is asymptotically stable, there is no trouble at all in extending the finite time results of the previous section to the infinite-time case. The limiting solution of the relevant Riccati equation exists, the optimal controller is time invariant, and the resulting system consists of a standard time-invariant optimal regulator with an external input. The complete observability condition ensures that the system is asymptotically stable. For this special case, then, the objectives set out at the beginning of this section have been achieved. Problem 11.3-1 asks that the details for this case be developed.

We consider now the case that arises when the desired trajectory $\tilde{y}$ is not asymptotically stable. Without loss of generality, we assume that $\tilde{y}$ is the output of the completely observable system

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \qquad \begin{bmatrix} z_1(t_0) \\ z_2(t_0) \end{bmatrix} \text{ given} \tag{11.3-7}$$

$$\tilde{y} = C_1'z_1 + C_2'z_2 \tag{11.3-8}$$

where the real parts of all the eigenvalues of A_1 are nonnegative and the real parts of all the eigenvalues of A_2 are negative. In other words, none of the states z_1 are asymptotically stable and all the states z_2 are asymptotically stable.

To consider this more general servo problem as a regulator problem, we introduce the following assumption.

ASSUMPTION 11.3-1. The nonasymptotically stable contribution to $\tilde{y}$—viz., $C_1'z_1$, is a zero-input response of the plant (11.3-1), (11.3-2). (*Note:* This implies that the eigenvalues of F include the eigenvalues of A_1.)

Assumption 11.3-1 implies that $\tilde{y}$ can be regarded as the output of the following system:

$$\begin{bmatrix} \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \begin{bmatrix} A_2 & 0 \\ 0 & F \end{bmatrix} \begin{bmatrix} z_2 \\ z_3 \end{bmatrix} \qquad \begin{bmatrix} z_2(t_0) \\ z_3(t_0) \end{bmatrix} \text{ given} \tag{11.3-9}$$

$$\tilde{y} = H'z_3 + C_2'z_2, \tag{11.3-10}$$

where $z_3(t_0)$ is appropriately chosen. This, in turn, means that the servo problem is once again a regulator problem where the plant

$$\begin{bmatrix} \dot{x} - \dot{z}_3 \\ z_2 \end{bmatrix} = \begin{bmatrix} F & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} x - z_3 \\ z_2 \end{bmatrix} + \begin{bmatrix} G \\ 0 \end{bmatrix} u \tag{11.3-11}$$

is to be controlled to minimize the index

$$V(x(t_0), z_3(t_0), z_2(t_0), u(\cdot))$$
$$= \lim_{T\to\infty} \int_0^T \left\{ u'Ru + \begin{bmatrix} x - z_3 \\ z_2 \end{bmatrix}' \begin{bmatrix} H \\ -C_2 \end{bmatrix} Q \begin{bmatrix} H \\ -C_2 \end{bmatrix}' \begin{bmatrix} x - z_3 \\ z_2 \end{bmatrix} \right\} dt.$$

Problem 11.3-2 asks that the details for this case be worked out.

Once again, the aims as stated at the beginning of this section have been achieved; in essence, the servo problem has been reduced to the regulator problem. The difficulty is simply that the key assumption, 11.3-1, is very restrictive. However, one important example of a case in which it is satisfied occurs when $\tilde{y}$ is a step function and the plant (11.3-1) and (11.3-2) is a single-input, single-output, "type 1" plant—i.e., the associated transfer function has a single pole at the origin (see Problem 11.3-4). But if in this example $\tilde{y}$ were a ramp input instead of a step function, then Assumption 11.3-1 would not be satisfied unless the plant were first augmented with an integrator at its input. This suggests, for the case when Assumption 11.3-1 is not satisfied by the original plant, that *compensation (either series or feedback) be included so that the resultant modified plant satisfies Assumption 11.3-1.*

As an example of the application of this principle, we shall now consider the case when a plant (11.3-1) and (11.3-2), with the property that F has no zero eigenvalues (i.e., F is nonsingular), is to follow as closely as possible a step function input $\tilde{y}$. We first augment the plant with integrators at the input and then consider the control of the augmented system

$$\begin{bmatrix} \dot{x} \\ \dot{u} \end{bmatrix} = \begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} \dot{u} \tag{11.3-12}$$

to minimize the index

$$V(x(t_0), u(t_0), \dot{u}(\cdot)) = \lim_{T\to\infty} \int_{t_0}^T \{\dot{u}'R\dot{u} + (y - \tilde{y})'Q(y - \tilde{y})\}\, dt. \tag{11.3-13}$$

Notice that this servo problem, for the case $\tilde{y} = 0$, reduces to the regulator

problem with derivative constraints given in the previous chapter. This servo problem can be solved by using the regulator theory of the last chapter; the details are requested in Problem 11.3-3. The optimal servo can be shown to consist of an optimal regulator having the form of those discussed in Chapter 10, with an external input signal. Once again, therefore, a method is given that achieves the objectives as stated in the introduction to this section.

So far, we have discussed servo problems completely reducible to regulator problems. It is straightforward to apply these ideas to the corresponding model-following problems. For the tracking problem case, if the desired trajectory is the output of a finite dimensional linear system, then the preceding results can be interpreted to give a solution to the tracking problem. Problem 11.3-4 is concerned with such a situation.

We now move on to consider what happens if we extend the finite-time result of the previous section to the infinite-time case without requiring that the resulting performance index be finite. Oddly enough, it turns out that results of engineering utility can still be obtained.

Direct extensions of finite-time results to the infinite-time case. For the system (11.3-1) and (11.3-2), performance index (11.3-3), and desired trajectory $\tilde{y}$ given by (11.3-4) and (11.3-5) (the servo problem), a control $\bar{u}$ may be found by taking the limits of the finite time results for the optimal control, provided, of course, that these limits exist. (Note that $\bar{u}$ is *not* necessarily optimal, but is still perhaps useful.) Assuming for the moment that these limits exist, we have that

$$\bar{u} = K'x + K_1'z \tag{11.3-14}$$

where

$$K' = -R^{-1}G'\bar{P} \quad K_1' = -R^{-1}G'\bar{P}_{21} \tag{11.3-15}$$

$$\bar{P} = \lim_{T\to\infty} P(t, T) \quad \bar{P}_{21} = \lim_{T\to\infty} P_{21}(t, T). \tag{11.3-16}$$

The equations for $P(\cdot, \cdot)$ and $P_{21}(\cdot, \cdot)$ are, as in the previous section, given from

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + HQH' \qquad P(T, T) = 0 \tag{11.3-17}$$

$$-\dot{P}_{21} = P_{21}F + A'P_{21} - P_{21}GR^{-1}G'P - CQH' \qquad P_{21}(T, T) = 0. \tag{11.3-18}$$

We have, from the infinite-time regulator theory of Chapter 3, Sec. 3.3, that since the pair $[F, G]$ is completely controllable, the limiting solution $\bar{P}$ of the Riccati equation (11.3-17) exists. Theory concerning the existence of $\bar{P}_{21}$ is not so straightforward.

To find conditions that ensure the existence of $\bar{P}_{21}$, we rearrange the

linear differential equation (11.3-18) using the relation (11.3-15) as follows:

$$-\dot{P}_{21} = P_{21}(F + GK') + A'P_{21} - P_{21}GR^{-1}G'(P - \bar{P}) + CQH'$$
$$P_{21}(T, T) = 0. \tag{11.3-19}$$

An expression for $(P - \bar{P})$ is given later in the book in Chapter 15, Sec. 15.2. This expression indicates that for any fixed t, $(P(t, T) - \bar{P})$ decays exponentially toward zero as T approaches infinity. Kreindler [1] has shown, using a result of [4], that if and only if $\bar{P}_{21}$ exists, the limit

$$\bar{\Pi}_{21} = \lim_{T \to \infty} \Pi_{21}(t, T) \tag{11.3-20}$$

also exists where $\Pi_{21}(\cdot, T)$ satisfies

$$-\dot{\Pi}_{21} = \Pi_{21}(F + GK') + A'\Pi_{21} - CQH' \qquad \Pi_{21}(T, T) = 0. \tag{11.3-21}$$

Moreover, $\bar{P}_{21} = \bar{\Pi}_{21}$. In contrast to (11.3-19), the existence of limiting solutions of (11.3-21) is easy to check. The solution of (11.3-21) may be verified to be

$$\Pi_{21}(t, T) = \int_t^T \exp[A'(t - \tau)]CQH' \exp[(F + GK')(t - \tau)]\, d\tau. \tag{11.3-22}$$

By inspections, we see, then, that $\bar{\Pi}_{21}$, and thus $\bar{P}_{21}$, *will exist if and only if the following assumption holds.*

ASSUMPTION 11.3-2. The sum of the real parts of any eigenvalue of A and any eigenvalue of $(F + GK')$ is negative.

With Assumption 11.3-2 holding, a method has been indirectly outlined for achieving a servo system with the property that it consists of an optimal, time-invariant regulator with an external input. Since the controllers for the system are time invariant, the objectives as stated in the introduction to this section are achieved. Observe, however, that no indication has been given concerning any performance index being minimized. Unless Assumption 11.3-1 holds, the infinite-time performance index is infinite. The only general optimal control interpretation of these results is that they are the limiting results of the finite-time optimal servo problem case. That is, they have properties very close to the optimal systems designed for a large terminal time T.

One important observation is that the results of Chapter 4 may be used to select a value of Q in the performance index to ensure that Assumption 11.3-2 is satisfied. By choosing a sufficiently large positive definite value of Q, the real parts of the eigenvalues of $(F - GK')$ may be guaranteed to be

less than any specified negative value. Clearly, the location of the eigenvalues of A will determine what this specified value is chosen to be. Problem 11.3-5 asks that the details for such a design be developed.

For the tracking problem, the control $\bar{u}$ is given from

$$\bar{u} = -R^{-1}G'(\bar{P}x + \bar{b}) \tag{11.3-23}$$

where

$$\bar{b} = \lim_{T\to\infty} b(t, T) \tag{11.3-24}$$

with $b(\cdot, \cdot)$ the solution of the linear differential equation

$$-\dot{b} = (F + GK')'b - HQ\tilde{y} \quad b(T, T) = 0. \tag{11.3-25}$$

For the particular case when $\tilde{y}$ is the output of a linear system (11.3-3) and (11.3-4) with Assumption 11.3-2 holding, then $\bar{b} = \bar{P}'_{21}z$, and $\bar{b}$ is clearly well defined. For other situations, to determine if $\bar{b}$ is well defined would require a separate investigation for each $\tilde{y}$, unless $\tilde{y}$ is asymptotically stable or is within a square integrable function of an output of a linear system.

An interesting application for the preceding theory arises when the system is single input, single output, and when $\tilde{y}$ is a step function. For this case, we have $\dot{x} = Fx + gu$, $y = h'x$, $Q = 1$, $R = \rho$, $k' = -\rho^{-1}g'\bar{P}$, $\bar{b} = [(F + gk')']^{-1}h\tilde{y}$, and $\bar{u} = -k'x - \rho^{-1}h'(F + gk')^{-1}g$. The following results may now be verified. The steady-state output is $\bar{y} = [h'(F + gk')^{-1}g]^2$; the steady-state control signal is $\bar{u} = [1 + k'(F + gk')^{-1}g][-h'(F + gK')^{-1}g]$; and the steady-state error is $\bar{e} = [h'(F + gk')^{-1}g]^2 - 1 = -[1 - k'(F + gk')^{-1}g]^2$. For the case when F is singular (i.e., has a zero eigenvalue), then $\det[I - (F + gk')^{-1}gk'] = \det[1 - k'(F + gk')^{-1}g] = 0$, and thus $h'(F + gk')^{-1}g = \pm\tilde{y}$, and $\bar{y} = \tilde{y}$ and $\bar{u} = 0$. An interesting property of this tracking system is that as ρ approaches zero, the poles of the closed-loop transfer function approach a Butterworth configuration, except for some poles equal in number to the number of zeros of the transfer function. These poles approach the zeros of the transfer function. Also, as ρ approaches zero, the bandwidth of the system and the control energy required increase. A discussion of this phenomenon is given in [5] and in Chapter 5, Sec. 5.4.

For the model-following problem on an infinite time interval, the servo results can be applied directly.

We conclude this section with a fairly lengthy example, drawn from [6] and [7]. The example illustrates the design of a controller for the lateral motion of a B–26 aircraft. The main idea is to provide control to cause the actual aircraft to perform similarly to a model; the way this qualitative notion is translated into quantitative terms will soon become clear.

The general equations governing the aircraft motion are the linear set

$$\begin{bmatrix} \dot{\phi} \\ \ddot{\phi} \\ \dot{\beta} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & L_p & L_\beta & L_r \\ \frac{g}{V} & 0 & Y_\beta & -1 \\ \frac{N_{\dot{\beta}} g}{V} & N_p & N_\beta + N_{\dot{\beta}} Y_\beta & N_r - N_{\dot{\beta}} \end{bmatrix} \begin{bmatrix} \phi \\ \dot{\phi} \\ \beta \\ r \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 \\ L_{\delta r} & L_{\delta a} \\ Y_{\delta r} & 0 \\ N_{\delta r} + N_{\dot{\beta}} Y_{\delta r} & N_{\delta a} \end{bmatrix} \begin{bmatrix} \delta_r \\ \delta_a \end{bmatrix}$$

In these equations, ϕ denotes the bank angle, β the sideslip angle, r the yaw rate, δ_r the rudder deflection, and δ_a the aileron deflection. Of course, we identify x with $[\phi \quad \dot{\phi} \quad \beta \quad r]'$, etc. The quantities L_p, etc., are fixed parameters associated with the aircraft. For the B–26, numerical values for the F and G matrices become

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -2.93 & -4.75 & -0.78 \\ 0.086 & 0 & -0.11 & -1.0 \\ 0 & -0.042 & 2.59 & -0.39 \end{bmatrix}$$

$$G = \begin{bmatrix} 0 & 1 \\ 0 & -3.91 \\ 0.035 & 0 \\ -2.53 & 31 \end{bmatrix}.$$

However, the dynamics represented by this arrangement are unsatisfactory. In particular, the zero-input responses are preferred to be like those of the model

$$\dot{z} = Az \qquad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -1 & -73.14 & 3.18 \\ 0.086 & 0 & -0.11 & -1 \\ 0.0086 & 0.086 & 8.95 & -0.49 \end{bmatrix}.$$

This model is derived by varying those parameters in the F matrix corresponding to aircraft parameters which could be physically varied. For this reason, the first and third rows of A are the same as the first and third row of F. The eigenvalues of A are

$$-1.065, \quad +0.00275, \quad -0.288 \pm j2.94.$$

Although one is unstable, the associated time constant is so large that the effect of this unstable mode can be cancelled by appropriate use of an external nonzero input.

To achieve actual performance resembling that of the model, we pose the quantitative problem of minimizing

$$\int_0^\infty [u'u + (x - z)'Q(x - z)]\, dt.$$

Thus, we are considering a servo problem that differs from the standard one only by the specializations $H = C = I$.

As we know, the optimal control becomes of the form

$$u = K'x + K_1'z$$

where K and K_1 are determined by the methods given previously. In practice, the control

$$u = K'x + K_1'z + u_{\text{ext}}$$

would be used, where u_{ext} denotes an externally imposed control. Likewise, in practice, the model equation $\dot{z} = Az$ would be replaced by

$$\dot{z} = Az + Gu_{\text{ext}}.$$

At this stage, we are faced with the problem of selecting the matrix Q in the performance index. From inspection of the performance index, it is immediately clear that the larger Q is, the better will be the following of the model by the plant. This is also suggested by the fact that large Q leads to some poles of the closed-loop system

$$\dot{x} = (F + GK')x$$

being well in the left half-plane—i.e., leads to the plant with feedback around it tending to respond fast to that component of the input, $K_1'z$, which arises from the model.

On the other hand, we recall that the larger Q is taken, the larger will be the entries of K, and, for the aircraft considered, it is necessary to restrict the *magnitude* of the entries of K to be less than those of

$$K_{\max} = \begin{bmatrix} 5 & 5 \\ 0.5 & 2 \\ 5 & 20 \\ 5 & 1 \end{bmatrix}$$

To begin with, a Q matrix of the form ρI can be tried. Either trial and error, or an approximate expression for the characteristic polynomial of $F + GK'$ obtained in [7], suggests that $Q = 5I$ is appropriate. This leads to

eigenvalues of $F + GK'$ which have the values

$$-0.99, \quad -1.3, \quad -5.13, \quad -9.14.$$

Larger Q leads to more negative values for the last two eigenvalues, whereas the first two do not vary a great deal as Q is increased. Comparison with the model eigenvalues suggests the possibility of further improvement, in view of the fact that the ratio of the nondominant eigenvalue of $F + GK'$ nearest to the origin to the eigenvalue of the model most remote from the origin is about 5. On the other hand, the gain matrix K associated with $Q = 5I$ is

$$\begin{bmatrix} -2.53 & -2.21 \\ -0.185 & -1.83 \\ 1.58 & 0.7 \\ -2.34 & -0.01 \end{bmatrix}$$

and at least one of the entries (the 2-2 one) is near its maximum permissible value in magnitude.

This suggests that some of the diagonal entries of Q should be varied. To discover which, one can use two techniques.

1. One can plot root loci for the roots of $F + GK'$, obtained by varying one q_{ii}. Variations causing most movement of the roots leftward and simultaneously retaining the constraints on K can be determined.
2. One can examine the error between the model state $z(t)$ and the plant state $x(t)$ obtained for several initial conditions, $z(0) = x(0) = [1 \ \ 0 \ \ 0 \ \ 0]'$, $[0 \ \ 1 \ \ 0 \ \ 0]'$, $[0 \ \ 0 \ \ 1 \ \ 0]'$ and $[0 \ \ 0 \ \ 0 \ \ 1]'$, for example, using the design resulting from $Q = 5I$. One can then adjust those diagonal entries of Q which weight those components of $(z - x)$ most in error.

Case 2 leads to the greatest errors being observed in $(z_2 - x_2)$ and $(z_4 - x_4)$. This suggests adjustment of q_{22} and/or q_{44}, and this is confirmed from (1). However, adjustment of q_{22} causes the 2–2 entry of K to exceed its maximum value. On the other hand, adjustment of q_{44}, from 5 to 20, proves satisfactory. The new eigenvalues of $F + GK'$ become

$$-0.908, \quad -0.66, \quad -9.09, \quad -11.2,$$

and the gain matrix K is

$$\begin{bmatrix} -0.201 & -2.23 \\ -0.185 & -1.83 \\ 1.42 & 0.164 \\ -4.42 & -0.264 \end{bmatrix}.$$

For completeness, we state the feedforward gain matrix K_1 associated

with the model states. For $Q = \text{diag}\,[5, 5, 5, 20]$, this is

$$K_1 = \begin{bmatrix} 0.101 & 2.045 \\ 0.344 & 2.172 \\ -2.153 & -1.54 \\ 5.61 & 2.42 \end{bmatrix}.$$

Although the model is unstable, the sum of any eigenvalue of the model matrix A and the matrix $F + GK'$ is negative, which as we know guarantees the existence of K_1.

Problem 11.3-1. Suppose we are given the time-invariant completely controllable plants (11.3-1) and (11.3-2), and the performance index (11.3-3). We are also given that the desired trajectory $\tilde{y}$ is the output of a linear, time-invariant, asymptotically stable system

$$\dot{z} = Az$$
$$\tilde{y} = C'z.$$

Apply regulator theory to determine a time-invariant optimal controller for this system.

Problem 11.3-2. Suppose we are given the time-invariant, completely controllable plant (11.3-1) and (11.3-2), and the performance index (11.3-3). We are also given that Assumption 11.3-1 is satisfied. Apply regulator theory to determine a time-invariant optimal controller for this system. How are the results modified for the case when the index is given by

$$V(x(t_0), u(\cdot)) = \lim_{T\to\infty} \int_{t_0}^{T} \{u'Ru + x'[I - H(H'H)^{-1}H']'Q_1[I - H(H'H)^{-1}H']x$$
$$+ (y - \tilde{y})'Q_2(y - \tilde{y})\, dt?$$

Problem 11.3-3. Suppose you are given the plant (11.3-1) and (11.3-2), where F has no zero eigenvalues, and also the performance index (11.3-13). Develop an optimal servo system for the case when the incoming signal $\tilde{y}$ is a step function. Where in the system can large sector nonlinearities be included without affecting the stability properties of the system?

Problem 11.3-4. Suppose you are given the system (11.3-1) and (11.3-2) where the output is to track $\tilde{y} =$ constant, minimizing the index (11.3-3).

1. Develop a tracking system for the case when F is singular (i.e., F has a zero eigenvalue) by reducing the problem to a regulator problem.
2. Develop a tracking system using tracking theory and discuss what happens when F is singular. Also, set up the tracking scheme so that the input is $\tilde{y}$ and there is no feedforward controller.

Problem 11.3-5. Suppose you are given a plant (11.3-1) and (11.3-2) and an incoming signal $\tilde{y}$ given from (11.3-4) and (11.3-5), where the real parts of the eigen-

values of A are all less than a positive constant α. Show how a finite-time performance index may be set up so that when minimized, the limiting optimal control, as the time interval of interest approaches infinity, will be finite. (*Hint:* Use the ideas of Chapter 4 to ensure that Assumption 11.3-2 is satisfied.)

Problem 11.3-6. For the infinite-time servo problem, where the performance index is always infinite for all controls, is it possible to set up the problem using an index

$$V(x(t_0), u(\cdot)) = \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{T} \{u'Ru + (y - \bar{y})'Q(y - \bar{y})\}\, dt$$

so that the difficulties associated with an infinite index do not occur? Discuss this case.

REFERENCES

[1] Athans, M., and P. L. Falb, *Optimal Control*, McGraw-Hill Book Company, New York, 1966.

[2] Kreindler, E., "On the Linear Optimal Servo Problem," *Intern. J. Control*, Vol. 9, No. 4, 1969, pp. 465–472.

[3] Kalman, R. E., "The Theory of Optimal Control and the Calculus of Variations," Ch. 16 of *Mathematical Optimization Techniques*, R. Bellman, (ed.) University of California Press, 1963.

[4] Bellman, R., *Stability Theory of Differential Equations*, McGraw-Hill Book Company, New York, 1953.

[5] Leake, R. J., "Return Difference Bode Diagram for Optimal System Design," *IEEE Trans. Auto. Control*, Vol. AC–10, No. 3, July 1965, pp. 342–344.

[6] Tyler, J. S., "The Characteristics of Model Following Systems as Synthesized by Optimal Control," *IEEE Trans. Auto. Control*, Vol. AC–9, No. 4, October 1964, pp. 485–498.

[7] Tyler, J. S., and F. B. Tuteur, "The Use of a Quadratic Performance Index to Design Multivariable Control Systems," *IEEE Trans. Auto. Control*, Vol. AC–11, No. 1, January 1966, p. 84–92.

CHAPTER 12

OPTIMAL REGULATORS WITH DUAL-MODE CONTROL

12.1 THE BANG-BANG MODE AND LINEAR MODE

In this section, a qualitative description is given of some optimal and suboptimal regulators with dual-mode control. The section serves as an introduction to the next section, which applies standard regulator theory to give useful results that may be used in the design of both optimal and suboptimal dual-mode regulators. The emphasis in the chapter is on the application of the results of earlier chapters to the problem of designing dual-mode regulators rather than on giving an exhaustive treatment of this subject for its own sake. Since an understanding of relay systems is essential to the development of the topic, a review of some of the relevant results of Chapter 6, Sec. 6.2, is now given.

A relay system having the form shown in Fig. 12.1-1 and having the closed-loop state equations

$$\dot{x} = Fx + g \operatorname{sgn}(k'x) \tag{12.1-1}$$

has two modes of operation. In one mode, the *bang-bang mode*, the input to the relay is either a fixed positive or fixed negative value, and any sign changes occur instantaneously. In the other mode, earlier termed the *chattering mode* but also known as the *sliding mode*, *singular mode*, or *linear mode*, the relay input is zero and the output chatters between its maximum and minimum output value with an average sliding value that could alternatively be realized by a linear state-feedback control law. This latter point is of

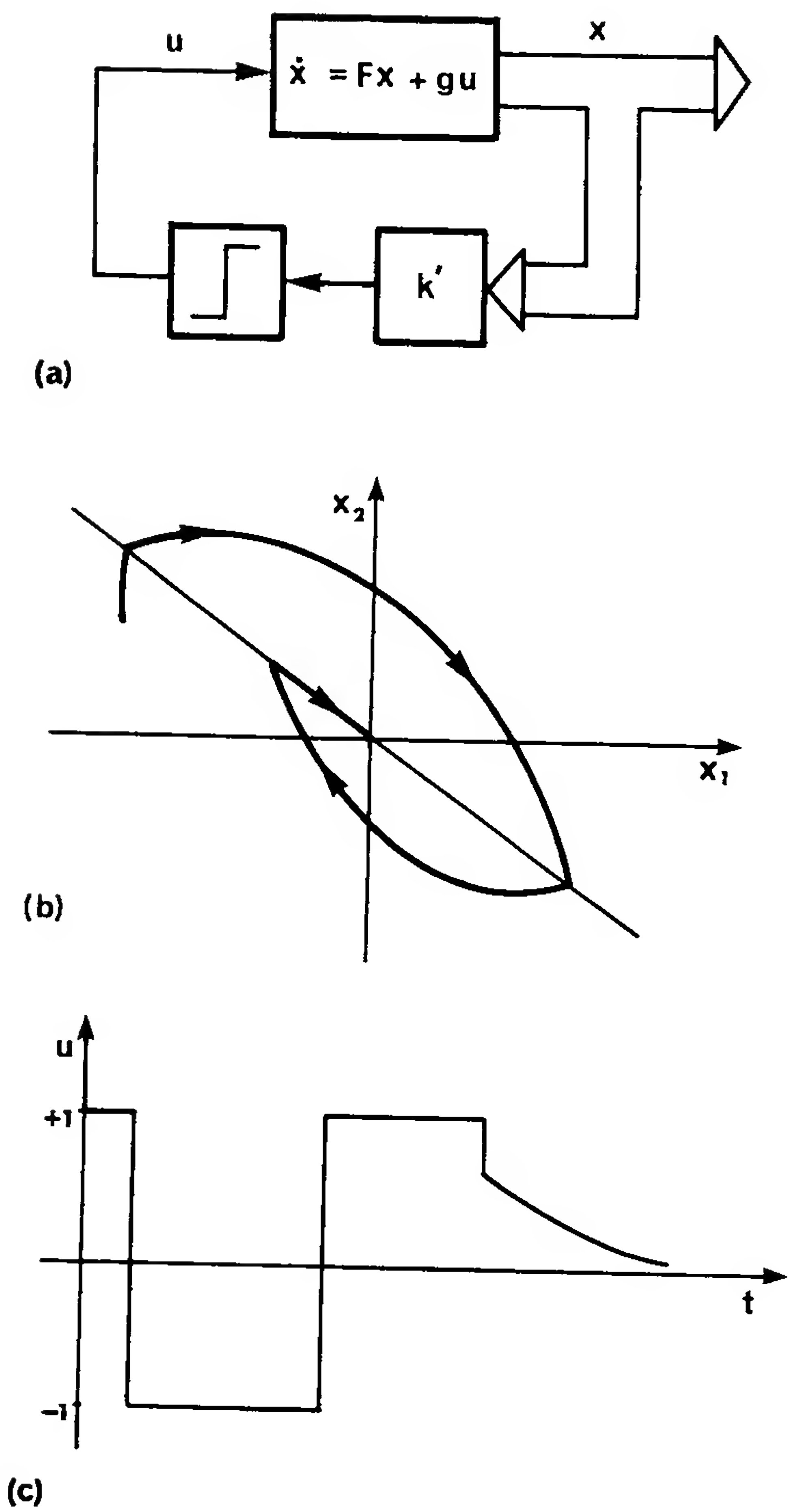


Fig. 12.1-1 Relay dual-mode control system.

particular interest to us in this chapter inasmuch as we are fundamentally concerned with linear control theory. Typically, for a large initial state, the regulator will operate first in its bang-bang mode and then, when the states are sufficiently small, it will enter the chattering mode and remain in this mode even on reaching the zero state. Figure 12.1-1 shows the form of such a relay system, a typical state trajectory and the corresponding control input. Note that the control indicated in Fig. 12.1-1(c) for the chattering mode is the average control rather than the true control.

For the chattering mode, the states obey the equations

$$k'x \equiv 0 \qquad |k'Fx| < k'g \tag{12.1-2}$$

and, if $k'g$ is nonzero,

$$\dot{x} = \left(I - \frac{gk'}{k'g}\right)Fx. \tag{12.1-3}$$

The bounded hyperplane defined by (12.1-2) will be termed the *singular bounded hyperplane* throughout the chapter. The *singular strip* is that portion of the singular bounded hyperplane with the property that solutions of (12.1-2) starting in the singular strip remain in the singular bounded hyperplane. For the case when x is a two-dimensional vector, the singular strip is the entire singular bounded hyperplane. For the case when x is a vector of size greater than 2, the relay system (12.1-1) may operate in the bang-bang mode and the chattering mode alternately until the state trajectory enters the singular strip. Once in the singular strip, of course, the trajectory will remain in the singular strip.

One quite significant property of the relay systems just described is that chattering mode performance is not at all affected by the introduction of a wide class of nonlinearities into the input transducers of the plant. That is, for the system

$$\dot{x} = Fx + g\beta \operatorname{sgn}(k'x), \tag{12.1-4}$$

where β is any positive time varying gain, the chattering mode trajectories are still given by (12.1-3), as in the case of system (12.1-1). Of course, the bang-bang mode trajectories will depend on β, but, as discussed in Chapter 6, the (nonglobal) stability properties of relay systems are usually dependent only on the stability properties of the chattering mode.

A rather obvious but significant property of the relay system is that the plant input is prevented from exceeding a specified upper limit. Of course, a saturation device rather than a relay at the plant input may be quite adequate in achieving the same purpose. Note that here, again, such a system operates dual mode. There is a *linear mode* and a *saturation mode*, as indicated in Fig. 12.1-2. This form of dual mode control leads on to a more general form, which we shall now discuss.

A dual-mode control system may be constructed with a switch that connects either of two controllers into the feedback loop at any one time. One controller implements a linear feedback law, and the other operates in a bang-bang mode, with the times of switching dependent on the location of the system trajectory in the state space. Figure 12.1-3 illustrates this more general form of control.

If an on-line computer is available, quite sophisticated forms of dual-mode control may be realized, and it may even be that in the case when the plant is highly nonlinear or the application is quite critical, only a sophisticated approach can achieve the desired results. As an example (later to be studed in some detail), we pose the problem of finding to a high order of

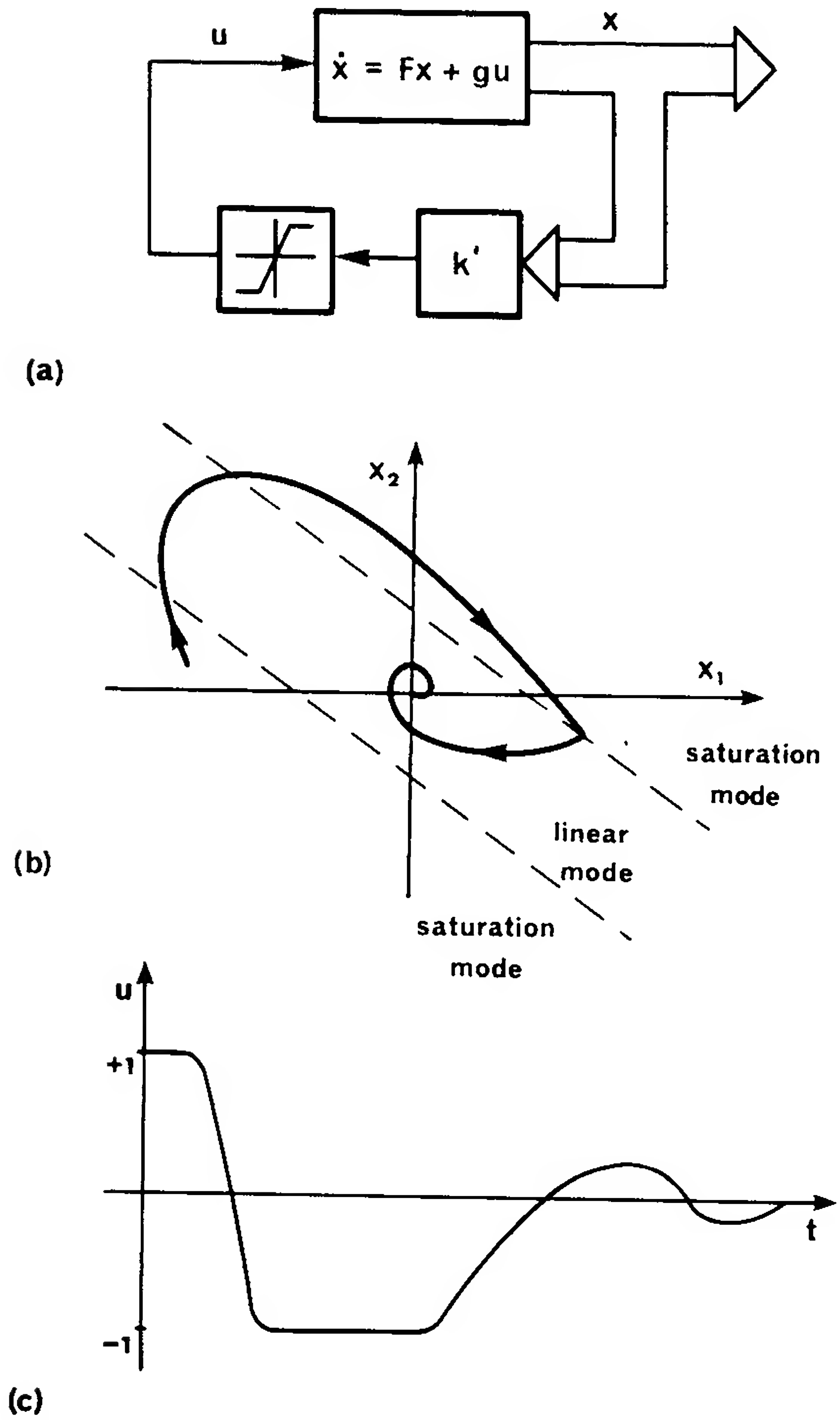


Fig. 12.1-2 Dual-mode regulator with saturation.

precision a control u for the plant

$$\dot{x} = Fx + gu \tag{12.1-5}$$

which will minimize the index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (x'Qx)\, dt \tag{12.1-6}$$

with Q nonnegative definite, subject to the inequality constraint

$$|u| \leq 1. \tag{12.1-7}$$

The application of optimal control theory to this *singular optimal control problem* yields the result that the optimal control is a bang-bang law over most of the state space with the switching surface being a hypersurface, which

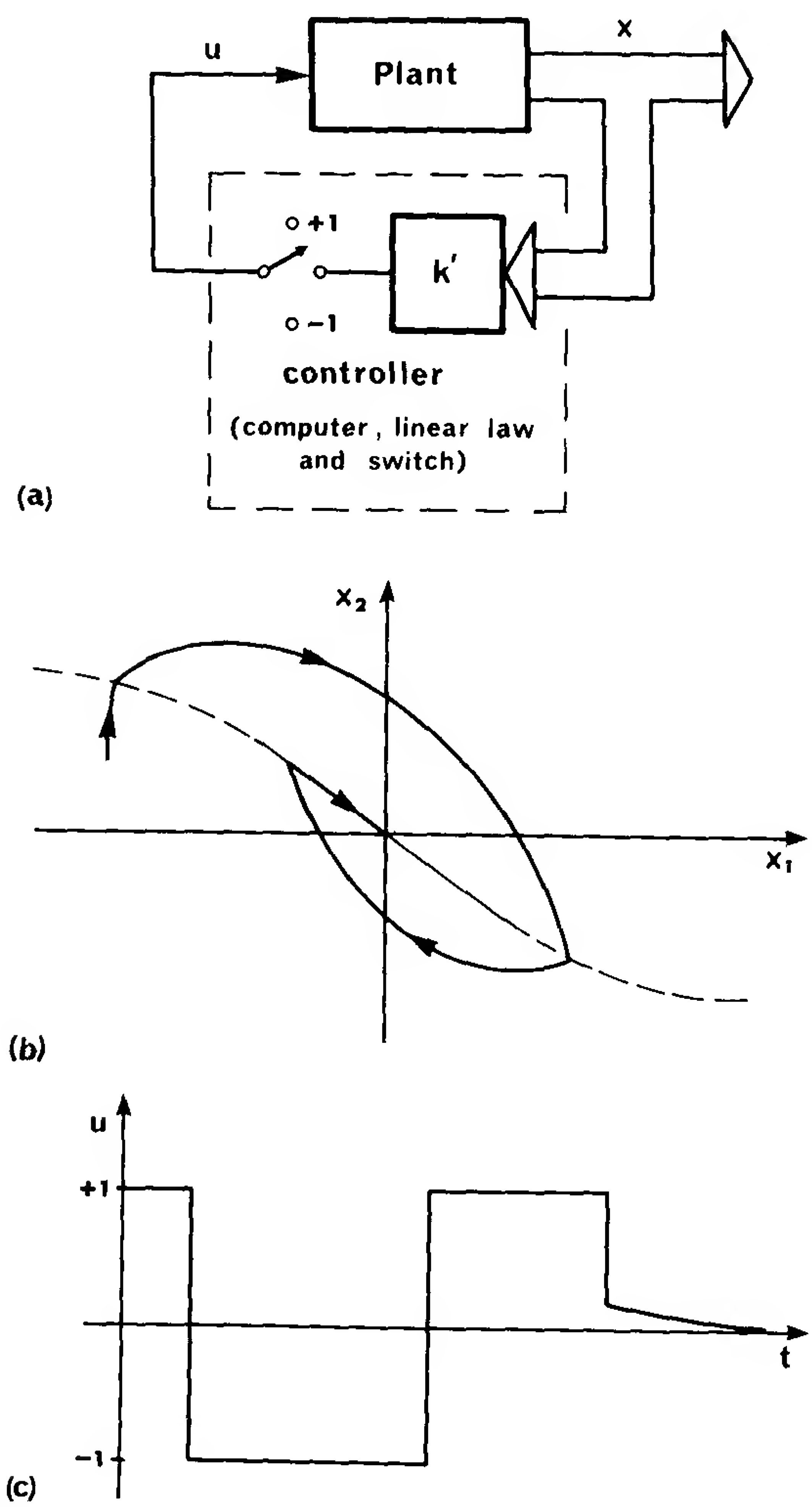


Fig. 12.1-3 Optimal dual-mode control.

is, in general, not linear. However, there is a certain portion of the hypersurface in the vicinity of the state space origin that is a bounded hyperplane (i.e., a linear hypersurface) known as the *singular strip*, for which the optimal control u^* satisfies $|u^*| < 1$. Once the state trajectory enters the singular strip, it remains in the singular strip.

To implement the optimal control law for the preceding problem requires a computer to store the characteristics of the hypersurface to decide when to switch from $u^* = +1$ to $u^* = -1$, and vice versa. But for the mode of operation for which $|u^*| < 1$, a computer is not needed. In fact, the control law is linear and is therefore readily realized as for the standard

regulator problem. It may also be realized by a relay in its chattering mode.

There is clearly a disadvantage in using the described form of dual-mode control—viz., an on-line computer is required. Unless a computer is available for other purposes as well, or unless it is vitally important to have precise optimal performance, this approach is probably not warranted, especially since by using other, more straightforward means nearly optimal performance may be achieved.

In fact, in the example, by using a simple relay controller, it is possible to achieve optimal performance for a region in the vicinity of the singular strip and nearly optimal performance outside this region. (The control law associated with the relay operating in its chattering mode is the same as the optimal, singular-strip control law.) When the relay controller is used, the optimal switching hypersurface is approximated by a hyperplane, which is an extension of the singular strip as indicated in Fig. 12.1-4. The singular

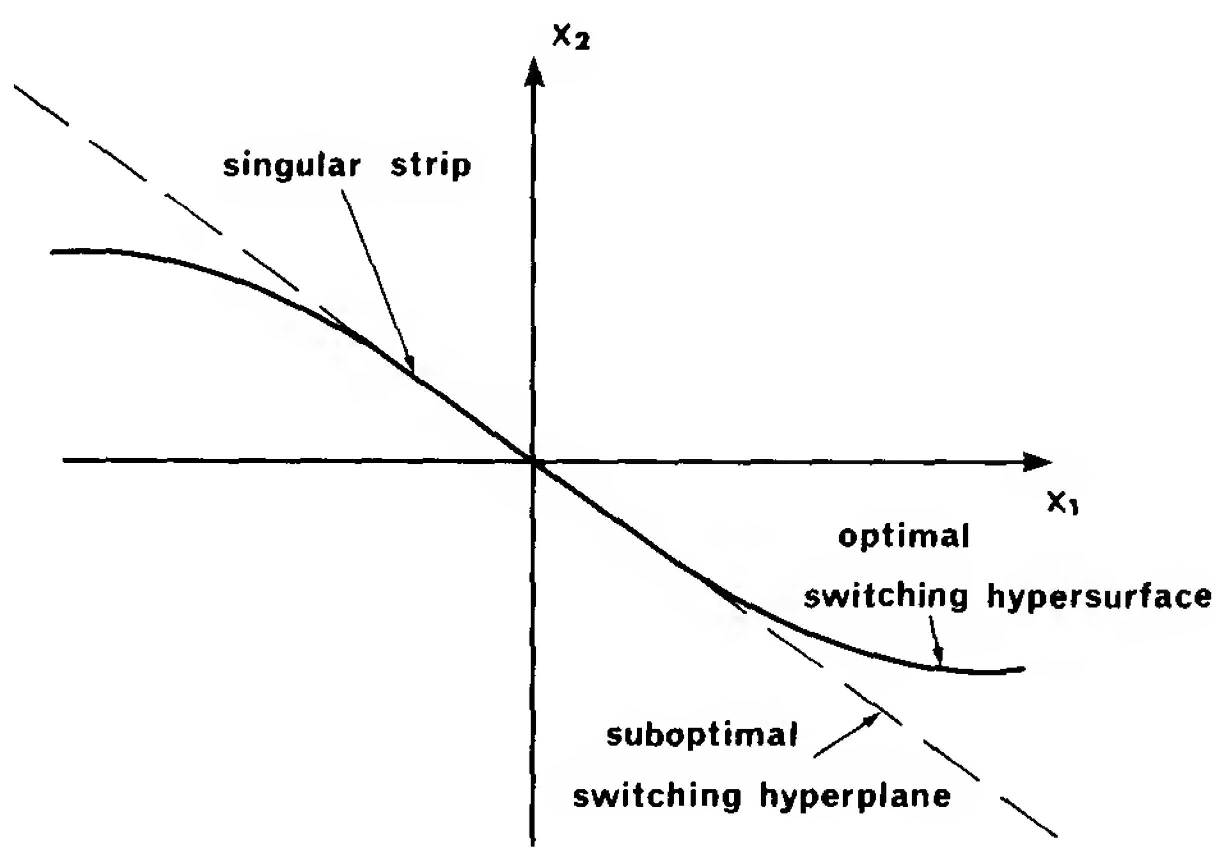


Fig. 12.1-4 Switching curves.

strip for the relay system (defined, we remember, as the set of points on the switching surface such that if chattering occurs at any such point, it occurs thereafter), will, of course, coincide with the singular strip for the optimal dual-mode system (i.e., the set of points such that the optimal control law takes on the linear form). Figure 12.1-5 indicates a region of the phase plane, which is the set of all initial conditions for the relay system that result in optimal state trajectories. For initial conditions outside this region, the state trajectories are suboptimal.

There are two important differences between the optimal system and the relay system that reflect in the state trajectories. The first difference is that the optimal switching surface is nonlinear, whereas the relay system switching surface is linear. Figure 12.1-6 illustrates the differences between the optimal

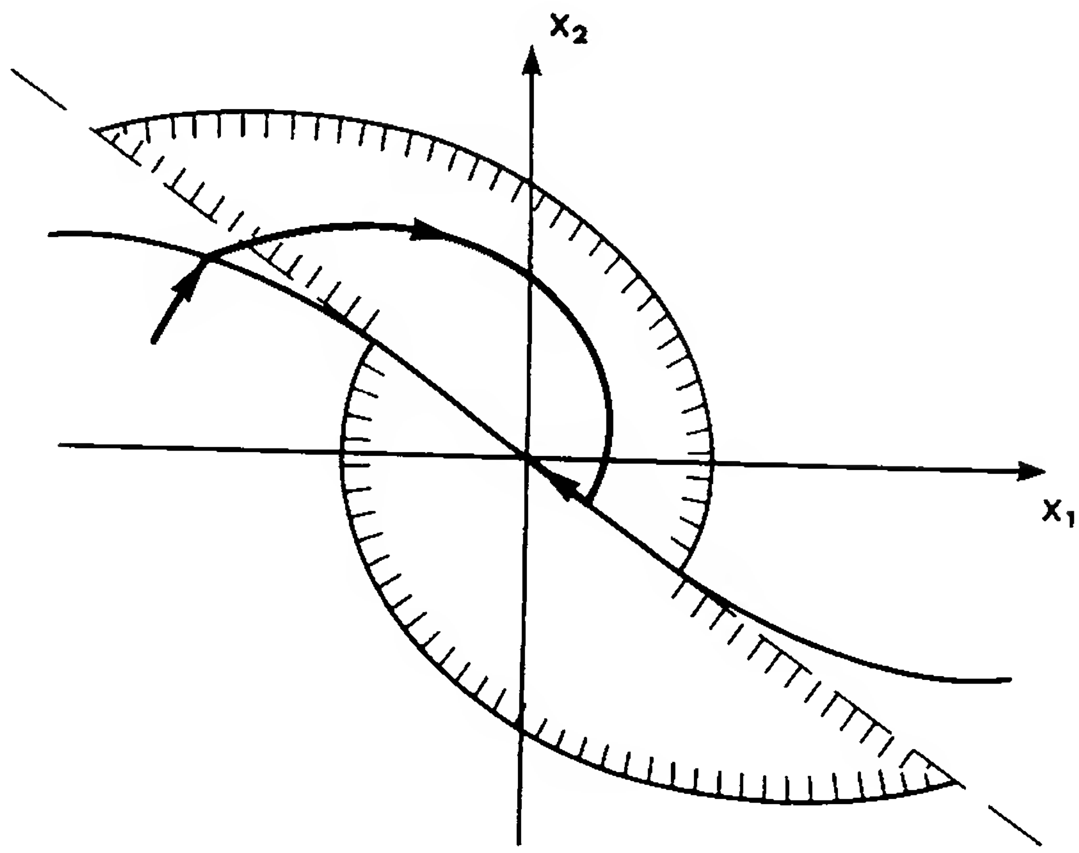


Fig. 12.1-5 State space region for which relay system is optimal.

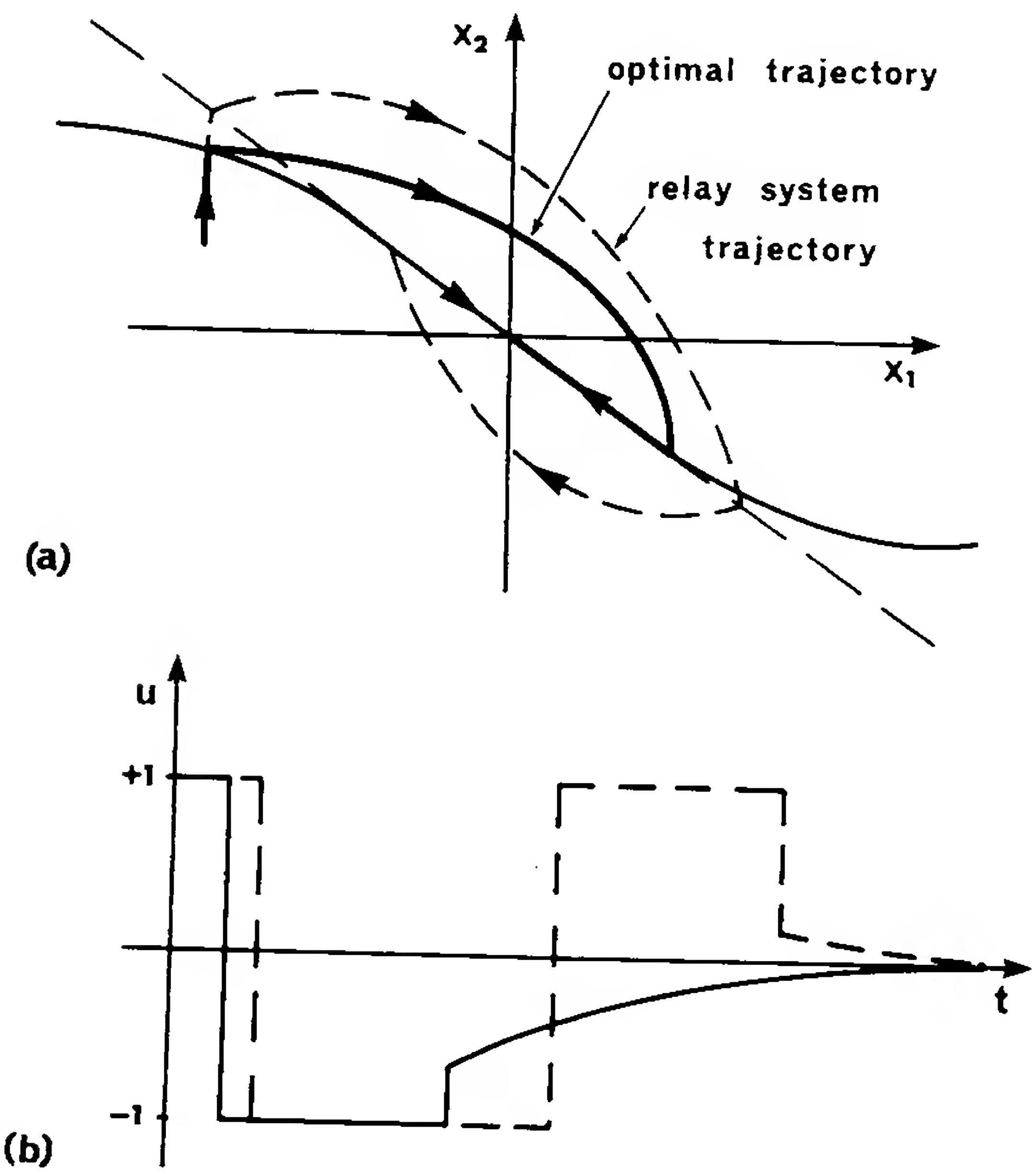


Fig. 12.1-6 Optimal and suboptimal trajectories.

trajectories and the suboptimal ones of the relay system due to this fact. Although the controls for the two cases are different, there is, in fact, very little difference between the performance indices $V = \int_{t_0}^{\infty} (x'Qx)\,dt$ for the two cases. The second difference is that chattering with the relay will occur in a singular bounded hyperplane, which is generally more extensive than the singular strip. The relay may switch from its chattering mode to the bang-

bang mode. This never happens for the optimal system unless there is an external disturbance. This difference is difficult to illustrate because it does not occur for a second-order system. But once again, the control may be significantly different without the performance indices differing very much.

As mentioned in Chapter 6, it might be that the engineer requires a control system with the same properties as the preceding suboptimal relay system, but for reliability or some similar consideration cannot use a relay in its chattering mode. For this case, a relay can be used until the commencement of chattering; then, a linear control law (say, $k_1'x$) can be switched in so as to achieve the same control law as the relay in its chattering mode. We observe that if the control law $k_1'x$ is changed to $(k_1' + \gamma k')x$ for some γ, then control is independent of γ as long as the trajectories lie within the singular bounded hyperplane, where the condition $k'x = 0$ is satisfied. Clearly, if there is some disturbance of the trajectories from the singular bounded hyperplane so that $k'x$ is no longer zero, then the effect of the control law $u = (k_1' + \gamma k')x$ will depend on the value of γ. It is therefore important to choose an appropriate value of γ to be used for any particular application, since, in practice, this law will be used in a region of the state space on either side of the singular bounded hyperplane.

Whatever the linear law used, it is necessary to consider how reversion to the bang-bang mode can be achieved. It might be that once the linear law has been switched in, it would be used even in the presence of a disturbance as long as the control satisfied $|u| < 1$. When $|u| = 1$, the relay would be switched in to bring the system back to the singular bounded hyperplane. On the other hand, a controller might monitor the magnitude of $k'x$, and when this exceeded a certain value, the relay with input $k'x$ would be used as the control law to return the state trajectory to the singular bounded hyperplane. Figure 12.1-7(a) and 12.1-7(b) indicates the alternative linear mode control arrangements. Passage from the linear mode to the bang-bang mode occurs at the boundaries of the shaded regions of the figures, and passage from the bang-bang mode to the linear mode occurs only on the singular bounded hyperplane.

The preceding remarks emphasize the scope for ingenuity in arriving at the best form of dual-mode control for any application.

A particularly attractive form of control law for the linear mode of operation of a dual-mode control scheme would be a law that minimized a quadratic index $\int_{t_0}^{\infty} (u'u + x'Q_1x)\,dt$, such as arises in the standard regulator problem. This would mean that in the linear mode of operation, the system would have all the desirable engineering properties associated with the standard regulator. We show in a later section that such a form of control law can be found by choosing the value of γ in the control law $u = (k_1' + \gamma k')x$ to be sufficiently large.

It might be argued that if the minimization problem posed in the first

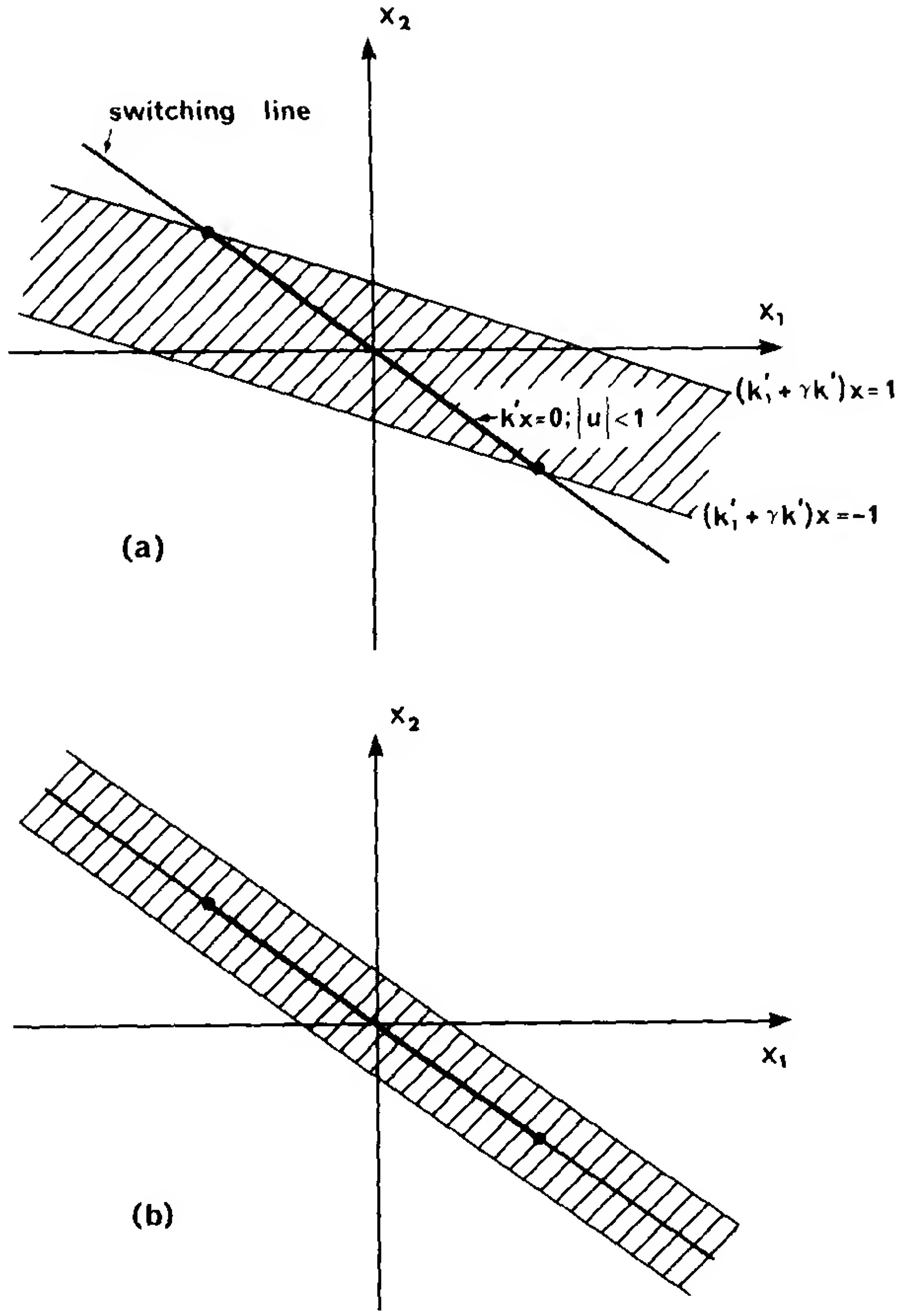


Fig. 12.1-7 Possible regions for linear-mode control after singular strip has been reached.

place is the minimization of an index

$$V = \int_{t_0}^{\infty} (u'u + x'Q_1 x)\, dt$$

for some nonnegative definite Q_1, subject to the basic system equation (12.1-5) and control magnitude constraint (12.1-7), then a system with the desirable properties discussed would be realized immediately. However, for this minimization problem, the optimal linear law is applicable in a region of the state space that is not confined to a hyperplane, and there is a region of the state space for which the optimal control law is nonlinear but not bang bang, as illustrated in Figure 12.1-8. For the case when Q_1 is chosen as ρQ for a constant ρ, and then ρ approaches infinity, the linear control law region

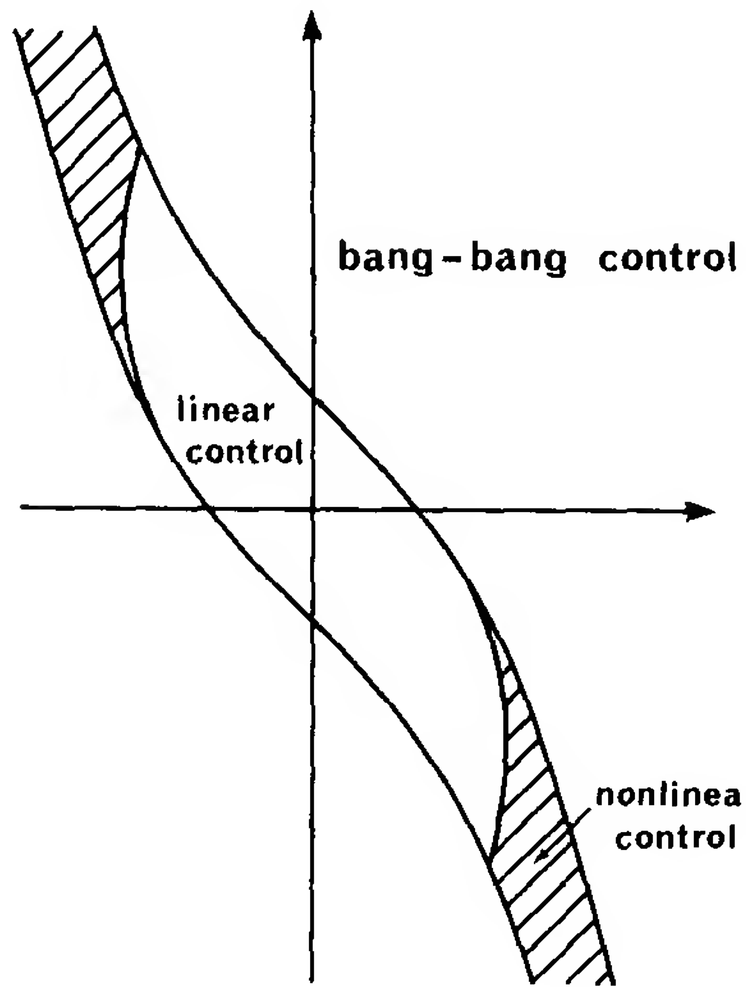


Fig. 12.1-8 Control regions resulting from minimizing a quadratic index with an upper bound on the control magnitude.

reduces to the singular strip of the original singular minimization problem, and the nonlinear region evanesces. But then the theory of the original minimization problem [Eqs. (12.1-5), (12.1-6), and (12.1-7)] is needed to study this case.

In the next section, we show that standard regulator theory can be applied to determine at least a part of the optimal solution for the problem of minimizing the index (12.1-2), subject to the constraints (12.1-5) and (12.1-7). In particular, the singular strip is described and the optimal control is found when the trajectories lie within this surface. That is, we solve for the *singular solutions* or *linear mode solutions* of the singular optimal control problem. Since the theory developing the optimal control law for the bang-bang mode cannot be readily applied without an on-line computer, it will not be discussed in this chapter. (For further details, consult reference [1].) As part justification, we note again the two points discussed earlier: Implementation of a suboptimal law using a relay or dual-mode control agreeing with the singular optimal law in the linear regime, but not the bang-bang optimal law in the bang-bang regime, leads to a value of the performance index not far different from the optimal value, and leads also to systems with pleasing engineering properties.

In the third section of the chapter, we consider the properties of dual-mode systems constructed by using the results of Sec. 12.2. Of particular interest is the ability to accommodate nonlinearities without introducing instability. A fourth section is included, which discusses the optimal control of a system subject to the constraint $|\dot{u}| \leq 1$. This is an application of the results of the previous sections together with the results of Chapter 10.

Much of the material of this chapter is taken from references [2] through [5]. For further background reading, references [6] through [10] are suggested.

12.2 LINEAR MODE SOLUTIONS TO THE SINGULAR OPTIMAL CONTROL PROBLEM

In this section, the regulator theory of the preceding chapters is applied to obtain the linear mode control laws or, equivalently, the singular solutions of what is known as the singular optimal regulator problem. A transformation is used to convert the problem of interest into a form of the standard regulator problem, and the standard regulator results are then interpreted so as to give the solution to the original problem. The results will be applied in subsequent sections to the design of dual-mode control systems consisting of a bang-bang mode and a linear mode. As a first step, we express the problem in mathematical terms.

The singular optimal regulator problem. We consider, as before, systems having state equation

$$\dot{x} = Fx + Gu \tag{12.2-1}$$

where F and G, if time varying, have continuous entries. As before, x is an n vector, the state, and u is an m vector, the input. Constraints are imposed on the system inputs, which may be expressed without loss of generality as simply

$$|u_i| \leq 1 \quad i = 1, 2, \ldots, m. \tag{12.2-2}$$

The performance index of interest is

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} (x'Qx)\, dt \tag{12.2-3}$$

for some finite $T \geq t_0$, where Q is a nonnegative definite symmetric matrix. The problem is to find the optimal control law $u^*(\cdot)$, which applied to the system (12.2-1) minimizes the index (12.2-3) subject to the inequality constraint (12.2-2). The infinite-time problem is defined by the limiting case $T \rightarrow \infty$.

For the preceding regulator problem, the optimal controller provides dual-mode control at each input; one mode, the bang-bang mode, operates when an input $u_i(t)$ takes on the values $+1$ or -1, whereas the other mode, the singular mode, occurs when $|u_i| < 1$. The latter is also referred to as a *linear mode*, because it may be realized by a linear control law, or as a *chattering mode*, because it may be realized by a relay with its output chattering

between $+1$ and -1. For this condition, the relay input is zero, and, as mentioned in the previous section, the terminology *singular mode* is used.

The preceding problem is an interesting mathematical one in its own right, but more important for us here is its engineering significance. We recall that for a standard regulator, the integrand of the performance index is $(x'Qx + u'Ru)$ with R positive definite. The term $u'Ru$ is included so that the optimal control u^* is bounded within reasonable limits. Here, the limits are set directly via the inequality constraint (12.2-2), and thus there is no need to include a $u'Ru$ term in the index for input limitation purposes. The fact that limits are set directly means that from an engineering point of view the singular optimal regulator problem is better posed than that for the standard regulator, provided, of course, that there is no restriction requiring the controller to be linear.

As a first step in solving the optimization problem, we consider the minimization of the index (12.2-3), subject to the system equation (12.2-1) *but not subject to the inequality constraints* (12.2-2). Once this problem has been solved, we shall discard those solutions that violate the constraints on u.

To solve for the singular solutions of this problem, we define new variables u_1 and x_1 through

$$\dot{u}_1 = u \tag{12.2-4}$$

$$x_1 = x - Gu_1 \tag{12.2-5}$$

and new parameter matrices

$$H = QG \quad R = G'QG \quad G_1 = FG - \dot{G}. \tag{12.2-6}$$

The variables u and x in (12.2-1) may be replaced by $\dot{u}_1$ and $x_1 + Gu_1$, respectively, which leads to

$$\dot{x}_1 = Fx_1 + G_1u_1. \tag{12.2-7}$$

Notice that (12.2-4) does not specify u_1 uniquely, but only to within a constant. Nevertheless, whatever u_1 is derived, Eq. (12.2-7) holds. Notice also that Eq. (12.2-1) follows from (12.2-7) and the definitions (12.2-4) and (12.2-5).

The performance index (12.2-3) may also be expressed in terms of x_1 and u_1 as follows:

$$\begin{aligned} V(x(t_0), u(\cdot), t_0) &= V(x_1(t_0), u_1(\cdot), t_0) \qquad \text{under (12.2-4) and (12.2-5)} \\ &= \int_{t_0}^{T} (x_1'Qx_1 + 2x_1'Hu_1 + u_1'Ru_1)\,dt \\ &= \int_{t_0}^{T} [x_1'(Q - HR^{-1}H')x_1 + (u_1 + R^{-1}H'x_1)'R \\ &\qquad \times (u_1 + R^{-1}H'x_1)]\,dt. \end{aligned} \tag{12.2-8}$$

We now see that the transformations (12.2-4) and (12.2-5) have converted the singular minimization problem into the quadratic regulator problem of Chap-

ter 3, Sec. 3.4, at least for the case when R is positive definite. For the case when $R = G'QG$ is not positive definite, the index will obviously not have the standard form, and further transformations are necessary. These cases are considered in Problem 12.2-1.

From the way the two minimization problems are related, we see immediately that under the converting relations (12.2-4) and (12.2-5), the minimization of (12.2-8), subject to the constraint (12.2-7), is precisely the same as the minimization of (12.2-3), subject to (12.2-1). As already noted, the minimization of (12.2-8), subject to the constraint (12.2-7), is essentially nothing other than a standard regulator problem.

Without further comment, we shall restrict attention to the case when $R = G'QG$ is positive definite. (Note that when Q is positive definite, this condition is automatically satisfied if the inputs are independent or if there is one input. Even in the case when Q is singular, there are many possible G matrices such that $R > 0$.)

We first claim that $(Q - HR^{-1}H')$ is nonnegative definite as a consequence of the nonnegativity of Q and the particular way in which H is defined. To see this, observe that $x'Qx$ is nonnegative for all x. Thus, using the transformations (12.2-4) and (12.2-5) as before, the term

$$[x_1'(Q - HR^{-1}H')x_1 + (u_1 + R^{-1}H'x_1)'R(u_1 + R^{-1}H'x_1)]$$

is nonnegative for all x_1 and u_1. For arbitrary x_1, setting $u_1 = -R^{-1}H'x_1$ leads to the conclusion that $x_1'(Q - HR^{-1}H')x_1$ is nonnegative and then that $(Q - HR^{-1}H')$ is nonnegative definite.

Since R is positive definite and $(Q - HR^{-1}H')$ is nonnegative definite, the standard regulator theory may be applied directly, as in Sec. 3.4, Chapter 3. For the finite-time case, the optimal control u_1^* is given by

$$u_1^* = K'x_1 \tag{12.2-9}$$

and the minimum index V^* is given as

$$V^*(x_1(t_0), t_0) = x_1'(t_0)P(t_0, T)x_1(t_0). \tag{12.2-10}$$

Here, K is given from

$$K' = -R^{-1}(G_1'P + H'), \tag{12.2-11}$$

and $P(\cdot, T)$ (where existence is guaranteed) is the solution of the Riccati differential equation

$$\begin{aligned}-\dot{P} = P(F - G_1R^{-1}H') + (F - G_1R^{-1}H')'P - PG_1R^{-1}G_1'P \\ + (Q - HR^{-1}H') \qquad P(T, T) = 0.\end{aligned} \tag{12.2-12}$$

Before interpreting these results in terms of the desired solution to the finite-time singular regulator problem, we shall perform some manipulations using the various equations and definitions to yield some simple relationships

between the various quantities hitherto defined. In particular, we shall establish in the order listed the relationships (valid for all t):

$$PG = 0 \qquad K'G = -I \qquad V^* = x'(t_0)P(t_0, T)x(t_0) \qquad K'x = 0.$$

Postmultiplying both sides of the Riccati equation (12.2-12) by the matrix G gives

$$-\dot{P}G = P(F - G_1R^{-1}H')G + (F - G_1R^{-1}H')'PG - PG_1R^{-1}G_1'PG$$
$$+ (Q - HR^{-1}H')G.$$

Applying the definitions (12.2-6) (viz., $R = H'G$, $H = QG$, and $G_1 = FG - \dot{G}$), gives immediately that

$$-\frac{d}{dt}(PG) = (F' - HR^{-1}G_1' - PG_1R^{-1}G_1')PG.$$

Now, since $P(T, T)G = 0$, the preceding differential equation in (PG) has the solution

$$PG = 0 \qquad \text{for all } t. \tag{12.2-13}$$

This is the first of the simple relationships.

Postmultiplying both sides of Eq. (12.2-11) by the matrix G gives that

$$K'G = -R^{-1}(G_1'PG + H'G).$$

Applying the result $PG = 0$ and the definitions (12.2-6)—in particular, $R = H'G$—yields the second simple result

$$K'G = -I, \qquad \text{together with } \dot{K}'G + K'\dot{G} = 0. \tag{12.2-14}$$

For the case when the control u_1 is, in fact, the optimal control u_1^*, then Eq. (12.2-5) becomes

$$x_1 = x - Gu_1^* \tag{12.2-15}$$

and the minimum index V^* given from (12.2-10) may be written as

$$V^*(x_1(t_0), t_0) = [x(t_0) - G(t_0)u_1^*(t_0)]'P(t_0, T)[x(t_0) - G(t_0)u_1^*(t_0)].$$

Since $PG = 0$ from (12.3-13), this becomes a function of only $x(t_0)$ and t_0 as follows:

$$V^*(x(t_0), t_0) = x'(t_0)P(t_0, T)x(t_0). \tag{12.2-16}$$

This is our third simple result. Finally, for the case when $u_1 = u_1^*$, we have that

$$K'x = K'x_1 + K'Gu_1^* = K'x_1 + K'GK'x_1.$$

Using the result that $K'G = -I$ [see (12.2-14)], we reduce this equation to our fourth simple result:

$$K'x = 0. \tag{12.2-17}$$

This result may seem somewhat paradoxical. A first glance at the earlier remarks would seem to indicate that, starting with an arbitrary $x(t_0)$, *not*

necessarily satisfying (12.2-17), one could set up an associated $x_1(t_0)$, solve the optimization problem associated with $x_1(t_0)$, and convert back to a solution of the optimization problem based on $x(t_0)$. The reason why this is not the case, and why a restriction such as (12.2-7) has to apply for all t and not even just for t_0, is as follows. Equation (12.2-15) requires that in passing from $x(\cdot)$ and $u(\cdot)$ to $x_1(\cdot)$ and $u_1(\cdot)$, the relation $x_1(t) + Gu_1^*(t) = x(t)$ should hold for all t.

This equation has two interpretations, one for $t = t_0$ and one for $t \geq t_0$. For the case of $t = t_0$, it implies that the optimization problem of minimizing (12.2-8) with respect to (12.2-7) for arbitrary $x_1(t_0)$ should be subjected to a side constraint on the value of the optimal control at time t_0—namely, $x_1(t_0) + G(t_0)u_1^*(t_0) = x(t_0)$, in order that its solution may be used to provide information concerning the original optimization problem. Without this constraint, we know that the optimum procedure is to set $u_1^*(t_0) = K'(t_0)x_1(t_0)$. Therefore, the modified optimization problem only yields a solution to the original optimization problem if, given $x(t_0)$, $x_1(t_0)$ can be chosen so that

$$x_1(t_0) + G(t_0)K'(t_0)x_1'(t_0) = x(t_0).$$

In view of the result $K'G = -I$, evidently a necessary condition on $x(t_0)$ is that $K'(t_0)x(t_0) = 0$. This is also a sufficient condition on $x(t_0)$, for if $x(t_0)$ satisfies $K'(t_0)x(t_0) = 0$, we can define $x_1(t_0) = x(t_0)$. Then $u_1^*(t_0) = K'(t_0)x_1(t_0) = K'(t_0)x(t_0) = 0$, and therefore the side constraint that $x_1(t_0) + G(t_0)u_1^*(t_0) = x(t_0)$ holds.

We now interpret the constraint equation $x_1(t) + Gu_1(t) = x(t)$ for $t > t_0$ assuming it holds for $t = t_0$. If $u_1^*(t)$ is the optimal control for the modified optimization problem, if (12.2-4) is used to set $u(t) = \dot{u}_1^*(t)$ as the control for the original problem (without consideration of its optimality or nonoptimality), and if $\dot{x} = Fx + Gu$ is used to determine $x(t)$, then it is also necessary for $x(t)$ to equal $x_1(t) + Gu_1(t)$. To see that this is indeed the case, observe that

$$\begin{aligned}\frac{d}{dt}[x(t) - x_1(t) - G(t)u_1^*(t)] &= \dot{x} - \dot{x}_1 - G\dot{u}_1^* - \dot{G}u_1^* \\ &= Fx + Gu^* - Fx_1 - G_1u_1^* - Gu^* - \dot{G}u_1 \\ &= F[x - x_1 - Gu_1^*].\end{aligned}$$

But since at $t = t_0$, $x(t_0) - x_1(t_0) - G(t_0)u_1^*(t_0) = 0$, we have $x(t) - x_1(t) - G(t)u_1^*(t) = 0$ for all t.

Consequently, $K'(t_0)x(t_0) = 0$ is a necessary and sufficient condition for the optimal control $u_1^*(t)$ of the modified problem to satisfy the constraint equation (12.2-15) for all $t \geq t_0$; this equation is, of course, necessary to draw conclusions concerning the first optimization problem for the second optimization problem. Note that (12.2-15) implies that $K'(t)x(t) = 0$ for all t. Therefore, $K'(t_0)x(t_0) = 0$ implies that $K'(t)x(t) = 0$ for all t.

We now return to the mainstream of the argument. Still considering the case when the control u_1 is optimal, i.e., when $u_1 = u_1^* = K'x_1$, we denote the corresponding control u as u^*, without claiming for the moment that u^* is optimal. From (12.2-4) and (12.2-9), we have that

$$u^* = \dot{u}_1^* = (\dot{\overline{K'x_1}}) = \dot{K}'x_1 + K'\dot{x}_1.$$

From the original system equation and the constraint equation (12.2-15), there follows

$$\begin{aligned} u^* &= \dot{K}'(x - Gu_1^*) + K'[F(x - Gu_1^*) + (FG - \dot{G})u_1^*] \\ &= \dot{K}'x + K'Fx - (\dot{K}'G + K'\dot{G})u_1^* \\ &= (K'F - \dot{K}')x \end{aligned} \tag{12.2-18}$$

where the final equality follows by use of (12.2-14).

We now claim that on applying the preceding control, viz., $u^* = (K'F + \dot{K}')x$ to the original system, provided that the initial state $x(t_0)$ is such that $K'x(t_0) = 0$, the performance index $\int_{t_0}^T (x'Qx)\,dt$ is minimized and has the value $V^* = x'(t_0)P(t_0)x(t_0)$. To see that this is so, we shall interpret the standard regulator results and the various relationships just developed. We have from the regulator results that applying the control $u_1^* = K'x_1$ to the system, $\dot{x}_1 = Fx_1 + G_1u_1$ minimizes the index (12.2-8). Equivalent to this minimization process for the case when $x(t_0)$ is chosen such that $K'(t_0)x(t_0) = 0$ is the application of the control $u^* = (K'F + \dot{K}')x$ to the system $\dot{x} = Fx + Gu$. To minimize the same index (12.2-8), subject to the constraints (12.2-4) and (12.2-5), is precisely the same as minimizing the index $\int_{t_0}^T (x'Qx)\,dt$, as we remarked earlier. Our claim is thereby established. An alternative argument that avoids use of the standard regulator theory is requested in Problem 12.2-2.

Of course, *by simply rejecting those solutions of the singular minimization problem which violate the constraints* (12.2-2) *at any time* $t \geq t_0$, *solutions that remain are the solutions to the singular optimal regulator problem posed at the beginning of this section, where we included the control magnitude constraint*. That is, for the ith input to operate in the singular mode rather than in the bang-bang mode, the constraints

$$|(K'Fx + \dot{K}'x)_i| < 1 \quad \text{and} \quad (K'x)_i = 0$$

must hold at all times.

We now recall the relay theory of Chapter 6, Sec. 6.2, to see that, provided the ith entry $u_i^*(t)$ of the control law u^* satisfies $|u_i^*(t)| < 1$ for all $t \geq t_0$, *the singular control law ith component may be realized by a relay in its chattering mode where the input to the relay is the ith entry of* $K'x$, written as $(K'x)_i$. In other words,

$$u_i^* = \operatorname{sgn}\,(K'x)_i. \tag{12.2-19}$$

As explained in the last section, (12.2-19) *also serves as a reasonable suboptimal control away from the singular strip*. Properties of the system resulting from using (12.2-19) will be discussed further in the next section.

The finite-time results just given can be readily extended to the limiting case as T approaches infinity, so long as we can establish the existence of the limit

$$\bar{P} = \lim_{T\to\infty} P(t, T). \tag{12.2-20}$$

Simply because both terms in the integrand of (12.2-8) are nonnegative, it follows that $x'(t)P(t, T)x(t)$ increases monotonically as T increases; all that is required to establish the existence of $\bar{P}$ is to show that $x'(t)P(t, T)x(t)$ has an upper bound for an arbitrary $x(t)$, which is independent of T. For the case when $[F, G_1]$ is completely controllable, the existence of the required upper bound is established by a direct application of the standard regulator results of Chapter 3, Sec. 3.1, but for our purposes, this condition is too restrictive.

We now claim that the condition that $[F, G]$ be completely controllable guarantees the existence of an upper bound on $x'(t)P(t, T)x(t)$, and thus that the limit $\bar{P}$ given by (12.2-20) exists. To see this, recall the existence of a control u_c, bounded but not necessarily by unity, which takes the system $\dot{x} = Fx + Gu$ from an initial state to the zero state in a finite time T_1 and is zero itself after T_1. Existence follows because $[F, G]$ is completely controllable. Clearly, by using the control u_c, we see that the index $V(x(t), u_c, t) = \int_t^T (x'Qx)\,dt$ is bounded above for all $T \geq t$ by some $V(\|x(t)\|)$. Now, Problem 12.2-2 establishes the following formula, valid for all T, $u(\cdot)$ and $x(t)$:

$$\int_t^T (x'Qx)\,dt = x'(t)P(t, T)x(t) + \int_t^T (x'KK'x)\,dt.$$

This implies that

$$x'(t)P(t, T)x(t) \leq V(x(t), u_c, t) < \bar{V} < \infty$$

with $\bar{V}$, of course, independent of T. If we let T approach infinity, it follows that

$$x'(t)\bar{P}(t)x(t) \leq \bar{V}(\|x(t)\|).$$

Since $x(t)$ is arbitrary, the desired result follows. The linear mode solutions to the singular optimal control problem are now summarized.

Linear mode solutions to the singular optimal regulator problem. Suppose we are given the completely controllable system

$$\dot{x} = Fx + Gu \tag{12.2-1}$$

and the performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^T (x'Qx)\,dt \tag{12.2-3}$$

for some nonnegative definite Q and some finite T. Suppose, also, that each input u_i is constrained by the inequality

$$|u_i| \leq 1. \tag{12.2-2}$$

Then the optimal control for the ith input of this system may be realized by a bang-bang control law (not given here) until the state trajectory reaches the singular strip, which is a subset of the singular bounded hyperplane

$$(K'x)_i = 0 \quad |(K'Fx + \dot{K}'x)_i| < 1 \tag{12.2-21}$$

where K' is given from (12.2-11) and (12.2-6) as

$$K' = -(G'QG)^{-1}[G'(FP + Q) - \dot{G}'P] \tag{12.2-22}$$

and P is the solution of the Riccati equation (12.2-12), which may be written

$$-\dot{P} = PF + F'P - KG'QGK' + Q \quad P(T, T) = 0. \tag{12.2-23}$$

Once the state trajectory is in the singular strip, it will remain there. (Recall that it is this property that defines the singular strip.) The control law u_i^* for the singular mode control may be realized by a controller gain $(K'F + \dot{K}')_i$ between the system states and input. An alternative realization is a controller gain $(K')_i$ followed by a relay that will operate in its chattering mode. The minimum value of the index $\int_{t_0}^{T} (x'Qx)\, dt$ for the case when $x(t_0)$ is in the singular strip is

$$V^*(x(t_0), t_0) = x'(t_0)P(t_0, T)x(t_0).$$

The results for the limiting case as T approaches infinity are the same as before, with $P(t, T)$ replaced by

$$\bar{P}(t) = \lim_{T \to \infty} P(t, T). \tag{12.2-20}$$

Figure 12.2-1 shows the two alternative realizations of the singular mode control for the single-input case. Lower-case letters are used to indicate vectors.

Problem 12.2-1. Given the completely controllable system $\dot{x} = Fx + gu$ and index $V = \int_{t_0}^{T} (x'Qx)\, dt$, with Q nonnegative definite, suppose that $g'Qg = 0$ and $(Fg - \dot{g})'Q(Fg - \dot{g}) \neq 0$. Find a control u^* to minimize the index. [*Hint:* Apply in succession the two transformations

$$x_1 = x - gu_1 \quad \dot{u}_1 = u$$

and

$$x_2 = x_1 - (Fg - \dot{g})u_2 \quad \dot{u}_2 = u_1$$

to convert the problem to a standard regulator problem.] See reference [3] for even more general results.

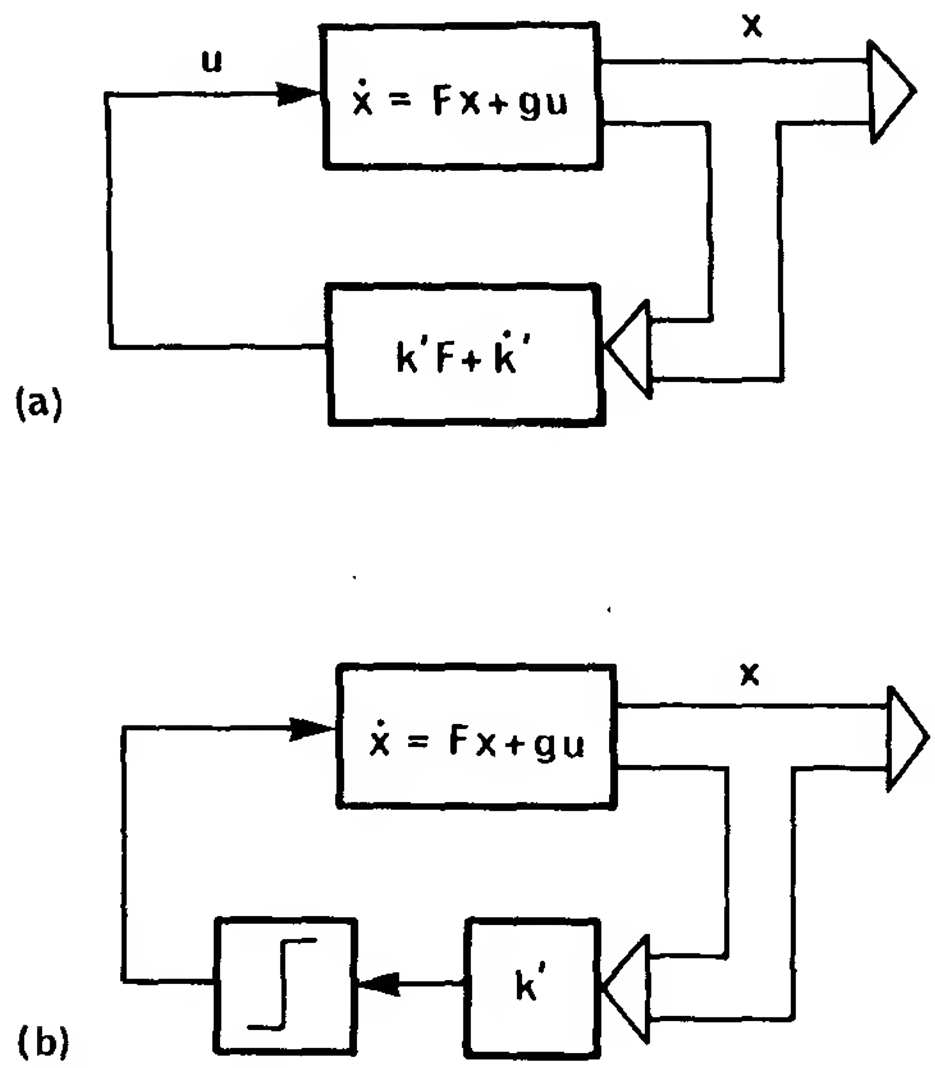


Fig. 12.2-1 Singular mode optimal control.

Problem 12.2-2. Show that the system equation $\dot{x} = Fx + Gu$ and Eqs. (12.2-11) through (12.2-13) imply the result

$$\int_{t_0}^{T} (x'Qx)\,dt = x'(t_0)P(t_0, T)x(t_0) + \int_{t_0}^{T} (x'KK'x)\,dt$$

independently of the control u. Use this result to verify the linear mode solutions to the singular regulator problem.

Problem 12.2-3. For the system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}1\\ -1\end{bmatrix}u$$

find the control that approximately minimizes

$$V = \int_0^T x_1^2\,dt$$

subject to $|u| \leq 1$. Verify that the appropriate controllability and observability conditions are satisfied. (*Hint:* In solving the Riccati equation, use the fact that

$$P\begin{bmatrix}1\\ -1\end{bmatrix} = 0.)$$

Problem 12.2-4. For the system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} + \begin{bmatrix}0\\ 0\\ 1\end{bmatrix}u$$

find the control that approximately minimizes the index

$$V = \int_0^T (x_1^2 + x_2^2 + x_3^2)\, dt$$

subject to $|u| \leq 1$. Observe that the Riccati equation is not difficult to solve when the result $P[0 \quad 0 \quad 1]' = 0$ is used.

12.3 PROPERTIES OF THE DUAL-MODE OPTIMAL REGULATOR

Throughout this book, we have adopted the philosophy that systems for which a quadratic-type performance index is minimized are of interest not so much because the index is at its minimum value but because such systems have ancillary properties desirable from the engineering point of view. Some of the properties of interest discussed in the previous chapters are degree of stability, gain margin, phase margin, system sensitivity to parameter variations, the ability to accommodate offset errors without affecting the equilibrium point (or desired operating point), and the ability to tolerate time delays and nonlinearities in the feedback loop while maintaining a prescribed degree of stability. Of course, optimality itself may be of direct engineering significance, as in the case of a system whose output is required to follow the output of a model while minimizing an error-squared type of performance index.

We now consider the "engineering" properties of systems derived by using the singular optimal control theory of the previous section. For simplicity, we consider single-input, time-invariant systems and performance indices measured over an infinite interval. That is, we consider the system

$$\dot{x} = Fx + gu \tag{12.3-1}$$

and performance index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{\infty} (x'Qx)\, dt \tag{12.3-2}$$

where it is assumed that Q is nonnegative definite and that $g'Qg$ is positive. Given these assumptions, we can, without loss of generality, normalize Q such that $g'Qg = 1$. The constraint inequality is simply

$$|u| \leq 1. \tag{12.3-3}$$

We recall from the previous section that the optimal control is bang-bang (i.e., $u = +1$ or $u = -1$) over the entire state space, except for the singular strip, which is a subset of the singular bounded hyperplane defined by

$$k'x = 0 \quad |k'Fx| < 1. \tag{12.3-4}$$

Here,

$$k' = -g'(F'\bar{P} + Q) \tag{12.3-5}$$

and

$$\bar{P} = \lim_{T\to\infty} P(t, T) = \lim_{t\to-\infty} P(t, T) \tag{12.3-6}$$

with $P(\cdot, T)$ the solution of

$$-\dot{P} = PF + F'P - (Q + PF)gg'(F'P + Q) + Q \qquad P(T, T) = 0. \tag{12.3-7}$$

On the singular strip, the optimal control law is linear and is given by

$$u^* = k'Fx \tag{12.3-8}$$

or by

$$u^* = k'(F + \gamma I)x \tag{12.3-9}$$

for any γ (since $k'x = 0$ on the singular strip). This law may be realized by a linear controller with gain $k'(F + \gamma I)$ for some γ. An alternative formulation for u^* on the singular strip is

$$u^* = \operatorname{sgn}(k'x). \tag{12.3-10}$$

The controller for this case is a linear controller of gain k' followed by a relay in its chattering mode.

To realize the optimal control law off the singular strip is difficult and may require an on-line computer. For this reason, we prefer to use one of the following realizations, not necessarily optimal, of the closed-loop system. In the first realization, the control law (12.3-10) is used for the entire state space. That is, the closed-loop system is

$$\dot{x} = Fx + g \operatorname{sgn}(k'x). \tag{12.3-11}$$

A second realization is to use system (12.3-11) until chattering commences and then to switch to the control law (12.3-9). This law may be used even if disturbed off the singular strip until, say, $|k'(F + \gamma I)x| = 1$ or $|k'x| = \delta$ for some positive δ chosen by the engineer. When either of these conditions is reached, the relay is switched back in the loop. (Recall the discussion in Sec. 12.1.) That is, we have a dual-mode control, where the closed-loop systems for the two modes are, respectively,

$$\dot{x} = Fx + g \operatorname{sgn}(k'x) \tag{12.3-12a}$$

and

$$\dot{x} = [F + gk'(F + \gamma I)]x. \tag{12.3-12b}$$

Since the system (12. 3-12b) will, in practice, be used off the singular strip, the value of γ should be chosen carefully. One result given in Chapter 6, Sec. 6.2, which is relevant in selecting γ, is that off the singular hyperplane

$k'x = 0$ the trajectories approach the hyperplane with a degree of stability γ. [To see this result, premultiply (12.3-12b) by k' and observe that this reduces to

$$\frac{d}{dt}(k'x) = -\gamma(k'x)$$

where we have used the relationship $k'g = -1$.] With $\gamma = 0$, trajectories off the hyperplane $k'x = 0$ will never approach the hyperplane, and the system (12.3-12b) is therefore not asymptotically stable.

We now discuss properties of the two dual-mode systems (12.3-11) and (12.3-12). The first property of interest is stability, a key aspect of which is the stability of both systems when $k'x = 0$ or, more precisely, when the trajectories are confined to the singular strip.

Stability on the singular strip. For the standard regulator problem, an observability condition is imposed as a sufficient condition for asymptotic stability of the closed-loop system. In the previous section, the minimization problem defined by (12.3-1) through (12.3-3) is transformed into a standard regulator problem in order to apply the known theory. In particular, the transformations

$$\dot{u}_1 = u \qquad x_1 = x - gu_1 \tag{12.3-13}$$

and definitions

$$h = Qg \qquad g_1 = Fg \qquad g'Qg = 1 \tag{12.3-14}$$

applied to (12.3-1) through (12.3-3) give the problem of minimizing the index

$$V(x_1(t_0), u_1(t_0), t_0) = \int_{t_0}^{\infty} [x_1'(Q - hh')x_1 + (u_1 + h'x_1)^2]\, dt \tag{12.3-15}$$

subject to the constraints

$$\dot{x}_1 = Fx_1 + g_1u_1 \qquad |\dot{u}_1| \leq 1. \tag{12.3-16}$$

For the case when the optimal control u_1^* for this problem without the constraint $|u_1| \leq 1$ still satisfies this inequality, it is given by

$$u_1^* = k'x_1 \tag{12.3-17}$$

with k' defined in the manner stated earlier. A sufficient condition for the closed-loop system $\dot{x}_1 = (F + g_1k')x_1$ to be asymptotically stable when the optimal control given by (12.3-17) is used is that the pair $[F - g_1h', D_1]$ be completely observable for any D_1 such that $D_1D_1' = (Q - hh')$. This is given from the regulator theory of Chapter 3, Sec. 3.3, applied directly to the preceding problem. With the closed-loop system asymptotically stable, u_1^*, given by (12.3-17), and x_1 decay asymptotically.

In the previous section, the transformations (12.3-13) are shown to be applicable for interpreting the optimal results for the preceding standard

regulator problem as results for the original minimization problem on the singular strip. Since the transformations (12.3-13) apply on the singular strip, a sufficient condition for $x = x_1 - gu_1$ to decay asymptotically is that x_1 and u_1 decay asymptotically. Expressing this condition in terms of the vectors and matrices of the original problem definition enables us to conclude that on the singular strip the system (12.3-1) with the control (12.3-12) applied is asymptotically stable with the following assumption.

ASSUMPTION 12.3-1. The pair $[F(I - gg'Q), D_1]$ is completely observable for any D_1 such that $D_1D_1' = Q(I - gg'Q)$.

That is, with Assumption 12.3-1 holding, the closed-loop system

$$\dot{x} = (I + gk')Fx \tag{12.3-18}$$

is asymptotically stable. If $x(t_0)$ satisfies

$$k'x(t_0) = 0 \quad |k'Fx(t_0)| < 1 \tag{12.3-19}$$

and if the condition $|k'Fx(t)| < 1$ remains satisfied for all t, both (12.3-11) and (12.3-12) are asymptotically stable. An alternative and simpler assumption now follows.

ASSUMPTION 12.3-2. The pair $[F_{CL}, D_2]$ is completely observable for any D_2 such that $D_2D_2' = Q$, where F_{CL} is the closed-loop system matrix $(I + gk')F$.

To see that this assumption is a sufficient condition for asymptotic stability on the singular strip, consider as a tentative Lyapunov function

$$V_1 = x'\bar{P}x. \tag{12.3-20}$$

Then

$$\begin{aligned}\dot{V}_1 &= 2x'\bar{P}\dot{x}\\ &= 2x'\bar{P}Fx + 2x'\bar{P}gu\\ &= -x'(Q - kk')x.\end{aligned} \tag{12.3-21}$$

The last equality is derived by using the result that $\bar{P}g = 0$ and the relationship $\bar{P}F + F'\bar{P} - kk' + Q = 0$. We observe that (12.3-21) holds irrespective of the control u. With our attention restricted to trajectories *on the singular strip* (i.e., $k'x = 0$), we note that the function V_1 is, in fact, the optimal performance index $V^*(x(t), t)$, for which

$$V^* = \int_t^\infty (x'Qx)\,dt. \tag{12.3-22}$$

Furthermore, on the singular strip, $k'x = 0$; therefore, $\dot{V}$ reduces to

$$\dot{V}^* = -x'Qx. \tag{12.3-23}$$

Clearly, if Q is positive definite, then V^* is a Lyapunov function, and trajec-

tories on the singular strip are asymptotically stable. However, in general, Q is but nonnegative definite, and thus V^* and $-\dot{V}^*$ are only guaranteed to be nonnegative. For V^* to be a Lyapunov function on the singular strip, we require first that V^* be strictly positive, and second, that $\dot{V}^*$ be not only nonpositive but also not identically zero on any system trajectory, confined, of course, to the singular strip (see the stability theorems of Chapter 3, Sec. 3.2). We now claim that Assumption 12.3-2 guarantees these results for the case when $x(\cdot)$ is the solution of (12.3-18) for some $x(t_0)$ satisfying (12.3-19). The reasoning is similar to that used in Chapter 3, Sec. 3.2, for the standard regulator stability results, and thus is only given in outline form here.

To see that V^* is positive on the singular strip with Assumption 12.3-2 holding, observe that if V^* were zero for some $x(t_0)$, then

$$0 = \int_{t_0}^{\infty} x'Qx\,dt = \int_{t_0}^{\infty} x'(t_0)\exp[F'_{CL}(t-t_0)]D_2D'_2 \times \exp[F_{CL}(t-t_0)]x(t_0)\,dt.$$

This equation implies that $D'_2\exp[F_{CL}(t-t_0)]x(t_0) = 0$ for a nonzero $x(t_0)$, and thus it contradicts Assumption 12.3-2.

To see that $\dot{V}^*$ is not identically zero along a system trajectory in the singular strip, suppose initially that it is identically zero. Then $x'Qx$ is identically zero, and thus V^* is identically zero, contradicting the previous result when Assumption 12.3-2 is in force.

The claim that Assumption 12.3-2 is a sufficient condition for stability on the singular strip has now been established. This means that under Assumption 12.3-2 the closed-loop system (12.3-11), when in its singular strip, is asymptotically stable. In fact, consider the introduction of nonlinearities β into the feedback loop as follows:

$$\dot{x} = Fx + g\beta[\operatorname{sgn}(k'x), t] \tag{12.3-24}$$

where β is a possibly time-varying nonlinearity, with the restrictions that $\beta(1, t)$ and $\beta(-1, t)$ are bounded,

$$\beta(1, t) \geq \beta_1 > 0; \quad \beta(-1, t) \leq \beta_2 < 0 \tag{12.3-25}$$

for all t, where β_1 and $-\beta_2$ are positive constants, then chattering still occurs on the singular strip in such a way as to leave the closed-loop system (12.3-18) unaltered. The stability results are thus unaffected (see Chapter 6, Sec. 6.2).

We also have that for the dual-mode system (12.3-12), once chattering of the relay commences for the system (12.3-12a) and the linear law is switched in resulting in closed-loop system equations (12.3-12b), then the system will behave in an asymptotically stable fashion, so long as the mode defined by (12.3-12b) continues to be the operating mode. However, as pointed out earlier, if there is a disturbance from these trajectories and $\gamma = 0$, then the system will not be asymptotically stable.

Problem 12.3-1 is concerned with establishing conditions that achieve a prescribed degree of stability α, say, *on* the singular strip. The problem of stability off the singular strip is now discussed.

Stability on and off the singular strip. In Chapter 6, Sec. 6.2, general results for relay systems are quoted. One of these results is that, provided the trajectories on the singular strip are asymptotically stable, then trajectories in the vicinity of the state-space origin are asymptotically stable. That is, local asymptotic stability is assured. The proof of the general result is not given in Chapter 6, but since, for the particular application of the closed-loop system

$$\dot{x} = Fx + g \operatorname{sgn}(k'x) \tag{12.3-11}$$

the result is readily proved, we now do so.

We consider the function

$$V_2 = x'\bar{P}x + (k'x) \operatorname{sgn}(k'x) \tag{12.3-26}$$

as a tentative Lyapunov function. Clearly, for the case when x is on the singular strip $k'x = 0$, then $V_2 = V_1 = V^*$, and, therefore, V_2 is positive definite. For the case $k'x \neq 0$, V_2 will be greater than or equal to $|k'x|$. Differentiating (12.3-26) yields

$$\begin{aligned} \dot{V}_2 &= -x'(Q - kk')x + k'[Fx + g \operatorname{sgn}(k'x)] \operatorname{sgn}(k'x) \\ &= -x'(Q - kk')x + k'Fx \operatorname{sgn}(k'x) - 1. \end{aligned} \tag{12.3-27}$$

For the second equality, the result $k'g = -1$ has been used. Observe that for sufficiently small x the term $\dot{V}_2$ is negative. This means that V_2 is a Lyapunov function for a sufficiently small region in the vicinity of the state-space origin. The local asymptotic stability result is thereby established.

It is also readily shown that the system given by (12.3-24) and (12.3-25) is locally asymptotically stable. That is, the introduction of nonlinearities of the form (12.3-25) into the system input transducers does not affect the preceding stability result.

We now ask if anything can be said about the *global asymptotic stability* of the system

$$\dot{x} = Fx + g \operatorname{sgn}(k'x). \tag{12.3-11}$$

One result is that (1) if the system $\dot{x} = Fx$ is asymptotically stable, and (2) if the real part of $(1 + jq\omega)k'(j\omega I - F)^{-1}g$ is positive for all ω and some nonnegative q, then the closed-loop system (12.3-11) is globally asymptotically stable. This stability result is discussed in Chapter 6, Sec. 6.2. Unfortunately, it does not help much here except in particular cases.

There is another result that could be mentioned, which is relevant to the problem. We claim that *if the system $\dot{x} = Fx$ is asymptotically stable,*

then the system that minimizes the performance index $\int_{t_0}^{\infty} (x'Qx)\,dt$ *will be globally asymptotically stable, assuming that the singular strip is asymptotically stable*. (We have already established that Assumption 12.3-2 guarantees the singular strip to be asymptotically stable.) To see that this result is so, we observe that the performance index is finite for a zero control and therefore must be finite for the optimal control. But from (12.3-20) and (12.3-21) (which are independent of the control u), we have

$$\int_{t_0}^{\infty} (x'Qx)\,dt = x'\bar{P}x + \int_{t_0}^{\infty} (x'kk'x)\,dt.$$

This means that with $\int_{t_0}^{\infty} (x'Qx)\,dt$ finite, then either $k'x = 0$ or $k'x$ approaches zero as t approaches infinity. However, the only way for $k'x$ to be zero or approximately zero for any length of time is for the trajectories to be in the vicinity of the singular strip. An examination of the function V_2 [see (12.3-26)] and its derivative [see (12.3-27)] shows that when Assumption 12.3-2 holds and when the trajectories are in the vicinity of the singular strip, V_2 is a Lyapunov function. Our claim is now established. Note again that this result is true in general only when the truly optimal switching surface is used, rather than the (suboptimal) hyperplane $k'x = 0$.

To establish global stability results for the dual-mode system (12.3-12) appears just as difficult as for the relay system (12.3-11). But *for the linear mode of operation, it is possible for the linear controller to be the same as that for a standard regulator*. This means that the linear mode of operation can have all the desirable properties associated with the regulator. This we now investigate.

Linear mode design for dual mode control. We present two approaches to the design of a linear mode controller for a dual-mode system. Both achieve a control law that is identical to one achieved using standard regulator theory, and both are based on the theory of the previous section. The idea of achieving the same controller as for a standard regulator is simply to guarantee the desirable properties associated with these regulators, such as a prescribed degree of stability, good phase margin and gain margin, etc.

The first approach requires a selection of the constant γ in the control $u = k'(F + \gamma I)x$. For any γ, the control will be optimal for the singular problem on the singular strip, whereas we can argue that for a suitably large γ, the control will also be optimal for a particular standard regulator minimization problem.

We know that for states on the singular strip, (provided Assumption 12.3-2 holds), the closed-loop system

$$\dot{x} = [F + gk'(F + \gamma I)]x \tag{12.3-12b}$$

is asymptotically stable. Moreover, the degree of stability orthogonal to the

singular strip is γ. In other words, $[F + gk'(F + \gamma I)]$ has $(n - 1)$ eigenvalues in the left half-plane and one eigenvalue at $s = -\gamma$. For γ sufficiently large, the eigenvalues will be the same as that for some standard optimal regulator design, as shown in Chapter 7, Sec. 7.4. Problem 12.3-2 asks for further details on this result.

An alternative approach to proving the same result is to consider as a tentative Lyapunov function for (12.3-12b) with input nonlinearities the following function:

$$V_3 = x'[2F'\bar{P}F + QF + F'Q + \gamma(Q + kk' + 2\gamma\bar{P})]x. \qquad (12.3\text{-}28)$$

Its derivative can be determined as

$$\dot{V}_3 = u^2(1 - 2\phi) - x'[F'(2\gamma\bar{P} - Q)F + \gamma^2 Q]x \qquad (12.3\text{-}29)$$

where ϕ denotes nonlinearities in the sector $[\frac{1}{2}, \infty)$ at the plant input transducer. For some suitably large γ, the function V_3 is, by inspection, a Lyapunov function (except for certain special cases). Because this is so, the theory on the inverse problem of optimal control, given in Chapter 7, yields that with a suitably large γ, the closed-loop system (12.3-12b) is optimal with respect to a standard quadratic performance index. Problem 12.3-3 asks that the details of this approach be worked out.

Problem 12.3-1. For the linear system $\dot{x} = Fx + gu$, $|u| \leq 1$, discuss the minimization of quadratic performance indices that achieve a prescribed degree of stability.

Problem 12.3-2. Using the terminology of this section, show that the matrix $[F + gk'(F + \gamma I)]$ has $(n - 1)$ eigenvalues in the left half-plane and one eigenvalue at $s = -\gamma$. What test could be applied to see if a particular γ is sufficiently large for the closed loop $\dot{x} = [F + gk'(F + \gamma I)]x$ to be optimal in the sense of Sec. 7.3?

Problem 12.3-3. Standard regulator theory yields an optimal control of the form $u^* = -g'\bar{P}_1x$, where the term $x'\bar{P}_1x$ is a Lyapunov function for the system $\dot{x} = (F - g\phi g'P_1)x$. (Recall that ϕ denotes a time-varying gain in the sector $[\frac{1}{2}, \infty]$.) The optimal control for the singular problem is $u^* = -g'(F'\bar{P} + Q)(F + \gamma I)x$. This suggests that we may be able to construct a Lyapunov function using the quadratic term $x'(F'\bar{P} + Q)(F + \gamma I)x$, associated with the system

$$\dot{x} = [F - g\phi g'(F'P + Q)(F + \gamma I)]x$$

for some $\gamma > 0$. Observe also that since $g'\bar{P} = 0$, the optimal control u^* could be written as $u^* = -g[(F'\bar{P} + Q)(F + \gamma I) + \frac{1}{2}\beta\bar{P}]x$ for some $\beta \geq 0$. All this suggests that we consider

$$V = x'[2(F'P + Q)(F + \gamma I) + \beta\bar{P}]x$$

as a tentative Lyapunov function. Investigate this case, using relationships such as $\bar{P}F + F'\bar{P} - kk' + Q = 0$ and $g'\bar{P} = 0$ to obtain for the case $\beta = 2\gamma^2$ the equations (12.3-28) and (12.3-29).

12.4 OPTIMAL REGULATOR WITH BOUNDS ON THE CONTROL DERIVATIVE

In this section, the theory of the previous sections is applied to solve the problem of controlling the completely controllable system

$$\dot{x} = Fx + gu \tag{12.4-1}$$

subject to the constraint

$$|\dot{u}| \leq 1. \tag{12.4-2}$$

The performance index of interest is the usual regulator index

$$V(x(t_0), u(\cdot), t_0) = \int_{t_0}^{T} (x'Qx + u^2)\, dt \tag{12.4-3}$$

where Q is nonnegative definite. We shall consider single-input, time-invariant plants, for simplicity. The physical significance of this problem is related to the regulator problem with derivative constraints of Chapter 10.

Clearly, the preceding minimization problem will be of interest for situations where rapid changes in the control input are to be avoided. It might be thought that making Q sufficiently small in the index (12.4-3) would ensure that there would be a small control, and therefore small rates of change of the control. In some situations, this may be true (although there is no immediate theoretical justification for it), but to choose an appropriate Q would depend on trial-and-error techniques, and result would probably be too conservative. With the constraint (12.4-2), the engineering limits are immediately taken into account in the design. The only disadvantage is an increase in controller complexity, as will become clear in the following development.

Using the same ideas as in Chapter 10, we define new variables

$$\hat{u} = \dot{u} \qquad \hat{x} = \begin{bmatrix} x \\ u \end{bmatrix}. \tag{12.4-4}$$

These definitions lead to

$$\dot{\hat{x}} = \hat{F}\hat{x} + \hat{g}\hat{u} \tag{12.4-5}$$

where

$$\hat{F} = \begin{bmatrix} F & g \\ 0 & 0 \end{bmatrix} \qquad \hat{g} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{12.4-6}$$

[Equation (12.4-5) represents a system consisting of the system of (12.4-1) augmented by an integrator at the input.] The performance index (12.4-3) is

$$V = \int_{t_0}^{\infty} (x'Qx + u^2)\, dt = \int_{t_0}^{\infty} (\hat{x}'\hat{Q}\hat{x})\, dt \tag{12.4-7}$$

where

$$\hat{Q} = \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix}. \tag{12.4-8}$$

The control variable constraint is now

$$|\hat{u}| \leq 1. \tag{12.4-9}$$

Application of the theory of the previous sections suggests that an effective suboptimal control is

$$\hat{u} = \operatorname{sgn}(\hat{k}'\hat{x}). \tag{12.4-10}$$

Here,

$$\hat{k}' = -(\hat{g}'\hat{Q}\hat{g})^{-1}\hat{g}'(\hat{F}\hat{P} + \hat{Q}) \tag{12.4-11}$$

and $\hat{P}(\cdot, T)$ is the solution of the Riccati equation, written using (12.4-11) as

$$-\dot{\hat{P}} = \hat{P}\hat{F} + \hat{F}'\hat{P} - \hat{k}(\hat{g}\hat{Q}\hat{g})^{-1}\hat{k}' + \hat{Q} \qquad \hat{P}(T, T) = 0. \tag{12.4-12}$$

The result for the augmented system (12.4-5) can now be interpreted as a result applying to the original system (12.4-1). Let us partition $\hat{P}$ as

$$\hat{P} = \begin{bmatrix} P_1 & P_2 \\ P_2' & P_3 \end{bmatrix}$$

using the same partitioning scheme as for $\hat{F}$ or $\hat{Q}$. Substituting all partitioned matrices into the equations (12.4-12) and (12.4-11) yields

$$\begin{aligned} -\dot{P}_1 &= P_1F + F'P_1 - P_1gg'P_1 + Q \\ P_2 &= 0 \\ P_3 &= 0 \\ \hat{k}' &= -g'P_1. \end{aligned} \tag{12.4-13}$$

Now, the control $\hat{u}$, or, equivalently, $\dot{u}$, is given as

$$\hat{u} = \dot{u} = \operatorname{sgn}(\hat{k}'\hat{x}) = \operatorname{sgn}(-g'P_1x - u).$$

This means that the differential equation for u is

$$\dot{u} = \operatorname{sgn}(-g'P_1x - u) \tag{12.4-14}$$

where P_1 is the solution of (12.4-13). This control scheme is shown in Fig. 12.4-1.

When the control $\hat{u}$ is on the singular strip $\hat{k}'\hat{x} = 0$, then we have that $-g'P_1x - u = 0$, or

$$u = -g'P_1x. \tag{12.4-15}$$

This is precisely the control law that results from the solution to the standard regulator problem of minimizing the index (12.4-3) for the original system (12.4-1). This result is expected, since the singular strip is that region where the control lies strictly within its constraints.

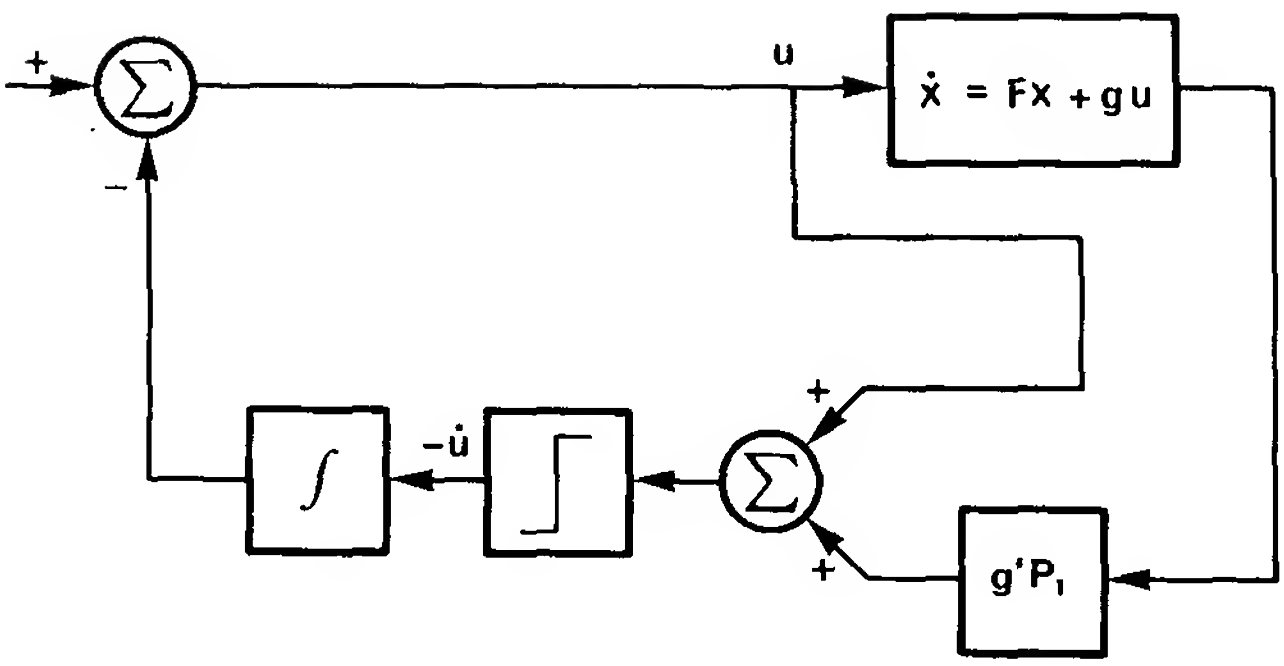


Fig. 12.4-1 Optimal regulator with $|\dot{u}| \leq 1$.

Further investigations of this result are left to the student in the problems.

Problem 12.4-1. Without verification, give a controller that could be used for the system (12.4-1) and index (12.4-3), where the control u must satisfy the constraint $|\ddot{u}| \leq 1$.

Problem 12.4-2. Can you think of a scheme which ensures that the constraints $|u| \leq 1$ and $|\dot{u}| \leq 1$ are satisfied?

Problem 12.4-3. Discuss how the system of Fig. 12.4-1 responds to a step input. Assume that the initial states of the plant and controller are zero. Redraw the system where it is required that the output track the input.

REFERENCES

[1] Sage, A. P., *Optimum Systems Control*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968.

[2] Moore, J. B., "A Note on a Singular Optimal Control Problem," *Automatica*, Vol. 5, No. 6, November 1969, pp. 857–858.

[3] Moylan, P. J., and J. B. Moore, "Generalizations of Singular Optimal Control Theory" (to be published).

[4] Moylan, P. J., and J. B. Moore, "Tolerance of Nonlinearities in Relay Systems," *Automatica*, Vol. 6, No. 2, March 1970, pp. 343–344.

[5] Moylan, P. J., and J. B. Moore, "An Optimal Regulator with Bounds on the Derivative of the Input," *Electron. Letters*, Vol. 5, No. 20, October 1969, pp. 502–503.

[6] Wonham, W. M., and C. D. Johnson, "Optimal Bang-Bang Control with Quadratic Performance Index," *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 86, March 1964, pp. 107–115.

[7] Webber, R. F., and R. W. Bass, "Simplified Algebraic Characterization of Optimal Singular Control for Autonomous Linear Plants," *Martin Marietta Corp. Res. Rep. R–67–6*, March 1967.

[8] Bass, R. W., and R. F. Webber, "On Synthesis of Optimal Bang-Bang Feedback Control Systems with Quadratic Performance Index," *Proc. Joint Auto. Control Conf.*, 1965, R. P. I., Rensselaer, N. Y., pp. 213–219.

[9] Rohrer, R. A., and M. Sobral, Jr., "Optimal Singular Solutions for Linear Multi-Input Systems," *Trans. ASME Ser. D: J. Basic Eng.*, June 1966, pp. 323–328.

[10] Sirisena, H. R., "Optimal Control of Saturating Linear Plants for Quadratic Performance Indices," *Intern. J. Control*, Vol. 8, No. 1, 1968, pp. 65–87.

CHAPTER 13

OPTIMAL LINEAR REGULATORS WITH CONTROLLER CONSTRAINTS

13.1 FORMULATION OF SPECIFIC OPTIMAL REGULATOR PROBLEMS

Virtually all the regulator results presented so far have been based on the assumptions that the system states, or estimates of these, are available for use in a control law and that there are no constraints other than linearity on the controller. In practice, however, further restrictions on the controller may be necessary. Therefore, in this chapter we develop some optimal linear regulator results with controller constraints. We refer to these problems as *specific optimal regulator problems*, *nonstandard regulator problems*, or *suboptimal regulator problems*.

Some of the results of this chapter first appeared in references [1] through [3]. The remainder of the results and some of the proofs of earlier results are new and have not, at the time of the writing of the text, appeared in the literature.

As in previous chapters, we assume we are given an n-dimensional linear system

$$\dot{x}(t) = F(t)x(t) + G(t)u(t) \tag{13.1-1}$$

$$y(t) = H'(t)x(t) \tag{13.1-2}$$

where $F(\cdot)$, $G(\cdot)$, and $H(\cdot)$ possess continuous entries. The simplest form of control law that we can consider is the nondynamic law

$$u(t) = K'(t)x(t). \tag{13.1-3}$$

Possible constraints on $K(\cdot)$ [other than the obvious one that $K(\cdot)$ be of appropriate dimension] that might be imposed in practice can be classified as follows:

1. Equality constraints on the entries of the controller gain matrix $K(\cdot)$, as, e.g., when only the outputs y are available for feedback. That is, state estimation is not permitted and the control u has the form
$$u = K_0'y = K_0'H'x.$$
Equivalently, K is constrained to have the form
$$K = HK_0$$
for some K_0 of appropriate dimension.
2. Inequality constraints on the entries of the controller gain matrix, as, e.g., when individual gain matrix elements k must satisfy
$$k_1 \leq k \leq k_2$$
for some constants k_1 and k_2. The reason for these constraints may be physical limitations involved in the construction of controllers, or the need to prevent saturation at the plant input.
3. Constraints that the controller gain have some prespecified functional form. We consider the special case of constant or piecewise constant gains. Clearly such constraints would be imposed for ease of practical implementation of the control law.

Of course, some or all of the constraints listed could be imposed simultaneously, if necessary.

In this chapter, we shall consider the optimal design of linear systems subject to constraints 1 and 3. The inequality constraints 2 can be handled (at least when constraints 3 are also imposed) using nonlinear programming techniques, such as are surveyed in [4]. These techniques will not be discussed since they are outside the scope of this book.

Also considered in this chapter are problems involving time-invariant plants arising when a dynamic controller is permitted between the plant output y and input u. We already know that if the dimension of the controller is not constrained in any way, the best that can be done in such a situation is to first construct a state estimator and then apply an optimal state feedback law using the state estimate in lieu of the inaccessible states. For the case when the controller dimension is constrained to be less than the dimension of the standard state estimator (see Chapter 8), then one approach is to estimate only some of the system states (or appropriate linear combina-

tions of the states) and use these estimates together with the system output in a nondynamic control law. The introduction of a (partial) state estimator then reduces the dynamic controller design problem to the nondynamic design problem. A second approach uses an artificial augmentation of the plant with integrators at its input; nondynamic feedback laws are determined for the augmented plant, which become dynamic laws for the original plant.

For the remainder of this section, a performance index will be considered, which is a natural extension of the usual quadratic index used in the standard regulator theory. With this index, an optimal control problem is formulated that incorporates controller constraints and is amenable to solution.

A performance index for regulator problems with controller constraints. As a starting point in our development, we recall the quadratic index associated with the standard regulator problem—viz.,

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} [u'(t)R(t)u(t) + x'(t)Q(t)x(t)]\, dt \tag{13.1-4}$$

with $R(t)$ positive definite symmetric and $Q(t)$ nonnegative definite symmetric for all t in the range $[t_0, T]$. To rewrite this index in terms of the present problem, we note that when control law (13.1-3) is used, the closed-loop system equations are

$$\dot{x}(t) = [F(t) + G(t)K'(t)]x(t). \tag{13.1-5}$$

The transition matrix $\Phi(t, \tau)$ associated with this system is defined from

$$\dot{\Phi}(t, \tau) = [F(t) + G(t)K'(t)]\Phi(t, \tau) \quad \Phi(\tau, \tau) = I. \tag{13.1-6}$$

Since $x(t) = \Phi(t, t_0)x(t_0)$ and $u(t) = K'(t)\Phi(t, t_0)x(t_0)$, the performance index (13.1-4) may be written simply as

$$V(x(t_0), u, T)|(u = K'x) = x'(t_0)P(t_0)x(t_0) \tag{13.1-7}$$

where

$$P(t_0) = \int_{t_0}^{T} \Phi'(t, t_0)[K(t)R(t)K'(t) + Q(t)]\Phi(t, t_0)\, dt. \tag{13.1-8}$$

The matrix $P(\cdot)$ is easily verified to be the solution of the linear equation

$$-\dot{P} = P(F + GK') + (F + GK')'P + (KRK' + Q) \tag{13.1-9}$$

with boundary condition $P(T) = 0$. See Appendix A for an outline discussion of such equations and their solution.

If we now impose constraints on $K(\cdot)$ and seek to find an optimum $K(\cdot)$ using the index (13.1-7) subject to these constraints, it will be generally true that for an arbitrary but fixed $x(t_0)$, there will be one (or possibly more) values of $K(\cdot)$ which will minimize (13.1-7). Unfortunately, the minimizing $K(\cdot)$ will probably vary from one $x(t_0)$ to another. However, for practical reasons, it is clearly desirable to have just one $K(\cdot)$ as the optimal feedback gain, irrespective of what $x(t_0)$ is. One way out of this dilemma is to assume

that the n vector $x(t_0)$ is a random variable, uniformly distributed on the surface of the n-dimensional unit sphere, and to seek the particular feedback gain $K(\cdot)$ minimizing the expected value, over all $x(t_0)$, of (13.1-7). That is, we define a *new performance index*

$$\bar{V}(K(\cdot), T) = E[x'(t_0)P(t_0)x(t_0)] \tag{13.1-10}$$

with the random variable $x(t_0)$ satisfying

$$E[x(t_0)x'(t_0)] = \frac{1}{n}I \tag{13.1-11}$$

where $E[\cdot]$ denotes expected value. {Problem 13.1-1 asks that an interpretation be given for the case when $E[x(t_0)x'(t_0)] = (1/n)A$ for some $A = A' > 0$.}

An alternative simpler expression for this index can be determined directly from (13.1-10) and (13.1-11) as follows:

$$E[x'Px] = \sum_{i,j} p_{ij}E[x_ix_j] = \frac{1}{n}\sum_i p_{ii} = \frac{1}{n}\text{tr}[P]$$

where x_i denotes the ith element of x, p_{ij} denotes the element in the ith row and jth column of P, and tr$[P]$ denotes the trace of P. Thus, the index

$$\bar{V}(K(\cdot), T) = \frac{1}{n}\text{tr}[P(t_0)] \tag{13.1-12}$$

is the new performance index with control variable $K(\cdot)$ [in the sense that $\bar{V}$ is a functional of $K(\cdot)$].

An alternative performance index to that given in (13.1-12) is $\bar{V} = \lambda_{\max}[P(t_0)]$. This gives the maximum value of $x'Px$ when x varies on the unit sphere, and in some situations it could be more useful than the index (13.1-12). However, we shall not consider this further in this chapter.

Let us summarize in one sentence the new minimization problem we have implicitly posed in the above arguments: Given matrices $F(\cdot)$, $G(\cdot)$, $Q(\cdot)$, and $R(\cdot)$, with the various restrictions assumed previously, find a matrix $K(\cdot)$ to minimize the index (13.1-10) with (13.1-11) holding, or, equivalently, the index (13.1-12), subject to whatever restrictions are imposed on $K(\cdot)$. This *nonstandard optimal control problem* has sometimes been referred to as a *suboptimal linear control problem* or a *specific linear control problem.*

When considering the limiting case as the terminal time T approaches infinity (the infinite-time case), we shall assume that the plant is time invariant and the control law is time invariant. {The optimal law for such a case may be time varying (see [5]).} With these assumptions, the optimal index is

$$\bar{V}(K) = E[x'(t_0)\bar{P}x(t_0)]$$

with

$$E[x(t_0)x'(t_0)] = \frac{1}{n}I$$

or, equivalently,

$$\bar{V}(K) = \frac{1}{n}\operatorname{tr}[P]$$

where

$$\bar{P} = \int_{t_0}^{\infty} \exp[(F + GK')'t](KRK' + Q)\exp[(F + GK')t]\,dt.$$

The matrix $\bar{P}$ is the solution of the linear equation

$$\bar{P}(F + GK') + (F + GK')'\bar{P} + (KRK' + Q) = 0.$$

Techniques for the solution of such equations are discussed in Appendix A.

For this infinite-time case, the optimum K must, of course, satisfy the further constraint that the closed-loop system

$$\dot{x} = (F + GK')x$$

be asymptotically stable. (Otherwise, the optimal closed-loop system will be of little practical utility. Also, $\bar{V}$ certainly will be infinite.) This raises the question as to whether there exists *any* constant control laws K (let alone an optimal one) that will satisfy the imposed constraints and the asymptotic stability constraint. Unfortunately, this existence question has as yet no straightforward answer, and the related problem of finding such a K, assuming one exists, has also not been solved in any convenient manner.

Of course, observability of $[F, D]$ for any D such that $DD' = Q$ is not sufficient to guarantee the desired stability, although this condition does guarantee that $\bar{P}$ is positive definite, provided that $F + GK'$ has all eigenvalues with negative real parts (see Problem 13.1-3).

Before we consider analytical results associated with minimizing the performance index just introduced, it is interesting to observe one immediate and simple link between the nonstandard optimal case and the standard optimal case. This is, that *if there are no constraints such as listed earlier in this section on the controller gain, then the controller gain* $K(\cdot)$, *which minimizes the index* (13.1-12) *associated with the nonstandard optimal problem, is the same gain resulting from the solution of the standard regulator problem with index* (13.1-4). To see this, let us assume that the optimum value of the index (13.1-12) is $(1/n)\operatorname{tr}[\tilde{P}]$ and the optimum value of the index (13.1-4) is $x'P^*x$. Assume the corresponding control laws are $\tilde{K}$ and K^*, and that they are different. Then, because $\tilde{K}$ is not optimal for the index (13.1-4), $x'\tilde{P}x \geq x'P^*x$ for all x, with strict inequality for some x. In other words, $\tilde{P} - P^*$ is nonnegative definite, and nonzero. Hence, $\operatorname{tr}(\tilde{P} - P^*) > 0$, and so $\operatorname{tr}(\tilde{P}) > \operatorname{tr}(P^*)$. This contradicts the optimality of $\tilde{K}$ for the index (13.1-12). Hence, $\tilde{K} = K^*$ and $\tilde{P} = P^*$.

When the optimal law is constrained, then, in general, $\tilde{K}$ will not be the same as K^*, the optimal law associated with the unconstrained problem. It appears, then, that the nonnegative quantity $\operatorname{tr}[\tilde{P}] - \operatorname{tr}[P^*]$ gives some meas-

ure of how "close" a nonstandard optimal regulator is to a standard one. This "closeness" concept is an interesting one that will be investigated in further detail in later sections.

Problem 13.1-1. What is the mathematical and practical significance of a performance index of the form

$$\bar{V} = x'(t_0)P(t_0)x(t_0)$$

where $E[x(t_0)x'(t_0)]$ is $(1/n)$ A for some positive definite symmetric A matrix—rather than the unit matrix? Show that for this case, $\bar{V} = \operatorname{tr}[AP]$.

Problem 13.1-2. A second approach to the problem of constructing constrained dynamic controllers relies on the augmentation ideas of Chapter 10. Discuss how the ideas of Chapter 10 could be used, and present a detailed problem formulation.

Problem 13.1-3. Show that if $(F + GK')$ has all eigenvalues with negative real parts, and if $[F, D]$ is completely observable for any D with $DD' = Q$, then the solution $\bar{P}$ of the following equation is positive definite:

$$\bar{P}(F + GK') + (F + GK')'\bar{P} + (KRK' + Q) = 0.$$

13.2 ANALYTICAL RESULTS AND COMPUTATIONAL ASPECTS

In this section, we derive analytical results that are useful in formulating algorithms to compute feedback gains for linear regulators with controller constraints. The algorithms themselves are not presented here except for the simplest and one of the most useful cases. For the other cases, the reader is referred to references [1] through [3]. Throughout the section, we consider cases arising when controllers are constrained to be nondynamic. However, as a later section shows, the results have application to the cases when estimates of some of the states (or linear combinations of the states) are available. When finite-time optimal control is considered, the matrices are permitted to be time varying, but when infinite-time results are derived toward the end of the section, the matrices are assumed constant. The infinite-time results have the most ready application, but there are stability difficulties, as will be discussed.

State feedback controllers with constant or piecewise constant gain elements. Let us consider the n-dimensional closed-loop system

$$\dot{x}(t) = [F(t) + G(t)K'(t)]x(t) \tag{13.2-1}$$

with $K(\cdot)$ either constant or piecewise constant. We seek to minimize the index

$$\bar{V}(K(\cdot), T) = \frac{1}{n}\operatorname{tr}[P(t_0)] \tag{13.2-2}$$

where $P(\cdot)$ is the solution of the linear equation

$$-\dot{P} = P(F + GK') + (F + GK')'P + (KRK' + Q) \tag{13.2-3}$$

with boundary condition $P(T) = 0$. That is,

$$P(t_0) = \int_{t_0}^{T} \Phi'(t, t_0)[K(t)R(t)K'(t) + Q(t)]\Phi(t, t_0)\, dt \tag{13.2-4}$$

where $\Phi(\cdot, \cdot)$ is the transition matrix associated with (13.2-1):

$$\dot{\Phi}(t, \tau) = [F(t) + G(t)K'(t)]\Phi(t, \tau) \qquad \Phi(\tau, \tau) = I. \tag{13.2-5}$$

We shall assume that the interval $[t_0, T]$ is divided into m fixed intervals (not necessarily equal) with K constant in each interval. We denote the value of the matrix K in the ith interval $[t_{i-1}, t_i]$ by K_i. For the moment, we assume no further constraints on K.

Clearly, necessary conditions for any set of K_i to be optimum are that

$$\frac{\partial \operatorname{tr}[P(t_0)]}{\partial K_i} = 0 \quad (i = 1, 2, \ldots, m). \tag{13.2-6}$$

(See Appendix A for the definition of differentiation of a scalar with respect to a matrix.)

We now seek an expression for $\partial \operatorname{tr}[P(t_0)]/\partial K_i$ in terms of the K_i and known quantities. The notation k_{ijl} will be used to denote the entry in the jth row and lth column of the matrix K_i, the notation e_i to denote a vector with all entries zero except the ith entry which is unity, and the notation α_i will denote a function with value 1 on $[t_{i-1} > t_i]$ and zero elsewhere.

Formally, then, differentiating (2-3) with respect to k, there obtains (on equating $\frac{\partial}{\partial t}\left(\frac{\partial P}{\partial k_{ijl}}\right)$ and $\partial \dot{P}/\partial k_{ijl}$)

$$\begin{aligned}-\frac{\partial}{\partial t}\left(\frac{\partial P}{\partial k_{ijl}}\right) &= \frac{\partial P}{\partial k_{ijl}}(F + GK') + (F + GK')'\frac{\partial P}{\partial k_{ijl}} \\ &\quad + \alpha_i e_j e_l'(RK' + G'P) + \alpha_i(KR + PG)e_l e_j'.\end{aligned}$$

From this equation and the boundary condition $\partial P(T)/\partial k_{ijl} = 0$, it follows that

$$\frac{\partial P(t_0)}{\partial k_{ijl}} = \int_{t_{i-1}}^{t_i} \Phi'(t, t_0)[e_j e_l'(RK' + G'P) + (KR + PG)e_l e_j']\Phi(t, t_0)\, dt.$$

Taking the trace of both sides, and using the well-known properties (see Appendix A), $\operatorname{tr}[A'(B + B')A] = 2\operatorname{tr}[A'B'A] = 2\operatorname{tr}[B'AA']$, gives the simplification

$$\frac{\partial \operatorname{tr}[P(t_0)]}{\partial k_{ijl}} = \int_{t_{i-1}}^{t_i} \operatorname{tr}[(KR + PG)e_l e_j' \Phi(t, t_0)\Phi'(t, t_0)]\, dt.$$

This result may be simplified even further by application of the property $\operatorname{tr}[ab'] = b'a$ to yield

$$\frac{\partial \operatorname{tr}[P(t_0)]}{\partial k_{ijl}} = \int_{t_{i-1}}^{t_i} e_j'\Phi(t, t_0)\Phi'(t, t_0)(KR + PG)e_l\, dt,$$

which, in turn, yields an expression for $\partial \operatorname{tr}[P(t_0)]/\partial K_i$—namely,

$$\frac{\partial \operatorname{tr}[P(t_0)]}{\partial K_i} = \int_{t_{i-1}}^{t_i} \Phi(t, t_0)\Phi'(t, t_0)[K_iR(t) + P(t)G(t)]\, dt. \tag{13.2-7}$$

Thus, the *necessary conditions for optimality are that*

$$\int_{t_{i-1}}^{t_i} \Phi(t, t_0)\Phi'(t, t_0)[K_iR(t) + P(t)G(t)]\, dt = 0. \tag{13.2-8}$$

This same result is derived in [1] via a different procedure.

Clearly, the necessary conditions and expressions for partial derivatives that we have just developed may be incorporated into algorithms for finding optimum values for the K_i (see [1]). However, as indicated in reference [1], the task of algorithm construction and application is a formidable one for both programmer and computer; therefore, further discussion of this aspect of the problem will not be included here.

For the case when K is constrained to be piecewise constant and there are a large number of time intervals used, the results are, as expected, reasonably close to the optimum $K(\cdot)$ derived assuming no constraints on K. On the other hand, when the number of intervals is small (say, less than five), the sequences of optimum K_i in some examples considered in [1] bear no apparent relation at all to the optimum $K(\cdot)$ with no constraints.

Output feedback controllers with constant or piecewise constant gains. We now ask: What are the necessary conditions for the same problem as just treated, with the additional constraint that the system be controlled by memoryless output feedback rather than memoryless state feedback? For this case, the control u has the form $u = K_0'y = K_0'H'x$ for some K_0 of appropriate dimension. That is,

$$K = HK_0. \tag{13.2-9}$$

Of course, $K_0(t) = K_{0i}$, a constant matrix, on the ith interval. Using (13.2-9) and repeating the same arguments as before, yields an expression for $\partial \operatorname{tr}P[t_0]/\partial K_{0i}$:

$$\frac{\partial \operatorname{tr}[P(t_0)]}{\partial K_{0i}} = \int_{t_{i-1}}^{t_i} H'(t)\Phi(t, t_0)\Phi'(t, t_0)[H(t)K_{0i}R(t) + P(t)G(t)]\, dt \tag{13.2-10}$$

where, of course,

$$\begin{aligned} -\dot{P} &= P(F + GK_0'H') + (F + GK_0'H')'P \\ &\quad + (HK_0RK_0'H' + Q) \end{aligned} \tag{13.2-11}$$

with boundary condition $P(T) = 0$, and

$$\dot{\Phi}(t, \tau) = [F(t) + G(t)K_0'(t)H'(t)]\Phi(t, \tau) \quad \Phi(\tau, \tau) = I. \qquad (13.2\text{-}12)$$

The *necessary conditions for optimality* are now

$$\int_{t_{i-1}}^{t_i} H'(t)\Phi(t, t_0)\Phi'(t, t_0)[H(t)K_{0i}R(t) + P(t)G(t)]\,dt = 0 \qquad i = 1, 2, \ldots, m. \qquad (13.2\text{-}13)$$

The computational problems associated with this condition for optimality are clearly only minor modifications of the ones where full state feedback is allowed.

Output feedback controllers with time-varying gains. We now relax the constancy or piecewise constancy constraint, but insist that $K(t) = H(t)K_0(t)$—i.e., demand output feedback. For this case, necessary conditions for optimality may be obtained from the piecewise-constant output-feedback case by a limiting operation. We let the time interval over which gains are to be constant be no larger than Δt, and then let Δt approach zero. Such a procedure can be justified by a theorem associated with the Ritz method for calculating optimal controls (see [6]). Condition (13.2-13) simply reduces to requiring that

$$H'(t)\Phi(t, t_0)\Phi'(t, t_0)[H(t)K_0(t)R(t) + P(t)G(t)] = 0$$

for all t in the range $[t_0, T]$. That is, K_0 must satisfy

$$K_0' = -R^{-1}G'P\Phi\Phi'H[H'\Phi\Phi'H]^{-1} \qquad (13.2\text{-}14)$$

where we have used Φ to denote $\Phi(t, t_0)$. (Problem 13.2-1 asks that the case when the inverse of $H'\Phi\Phi'H$ fails to exist should be considered.) Of course, P and Φ satisfy (13.2-11) and (13.2-12), respectively. *These necessary conditions are also sufficient conditions*, as may be seen by forming a Hamiltonian and showing that the Hamilton–Jacobi equation is satisfied. This is done in reference [2]. (See also Problem 13.2-2.)

It is interesting to observe that when $H = I$, corresponding to state-variable feedback, then K_0 in (13.2-14) satisfies

$$K_0' = -R^{-1}G'P$$

and P is given from

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \qquad P(T) = 0.$$

In other words, the now familiar *standard finite-time regulator results* are recovered.

Infinite-time results for output feedback controllers with constant gains. Let us consider the time-invariant, n-dimensional system

$$\dot{x} = Fx + Gu$$

$$y = H'x$$

where the control u is constrained to have the form

$$u = K_0'y = K_0'H'x$$

for some constant matrix K_0 of appropriate dimension. We seek to minimize the performance index

$$\bar{V}(HK_0) = \frac{1}{n}\operatorname{tr}[\bar{P}]$$

where $\bar{P}$ is the solution of the linear equation

$$\bar{P}(F + GK_0'H') + (F + GK_0'H')'\bar{P} + (HK_0RK_0'H' + Q) = 0$$

provided, of course, that the closed-loop system

$$\dot{x} = (F + GK_0'H')x$$

is asymptotically stable. (Otherwise, the index has no physical significance.)

The *necessary conditions* for optimality may be determined from the limiting case of the finite-time result (13.2-13) as

$$\int_0^\infty H' \exp[(F + GK_0'H')t] \exp[(F + GK_0'H')'t][HK_0R + \bar{P}G]\, dt = 0$$

or, more conveniently, as

$$H'M(HK_0R + \bar{P}G) = 0$$

where M is the solution of the linear equation

$$M(F + GK_0'H')' + (F + GK_0'H')M + I = 0.$$

[Of course, these results could be proved directly by using the same approach as that in the previous section. This is requested in Problem 13.2-3—(see also [3]).]

An alternative statement of the necessary condition for optimality is that K_0 satisfy

$$K_0' = -R^{-1}G'\bar{P}MH(H'MH)^{-1}.$$

[Compare this result with the time-varying result (13.2-14).] Using the notation * to indicate optimality, we now rewrite the necessary conditions for optimality as follows:

1. $F + GK_0^{*\prime}H'$ has all eigenvalues with negative real parts.
2. There exist positive definite symmetric matrices $\bar{P}$ and M such that the following three equations hold:

$$\bar{P}(F + GK_0^{*\prime}H') + (F + GK_0^{*\prime}H')'\bar{P} + HK_0^*RK_0^{*\prime}H' + Q = 0$$

$$M(F + GK_0^{*\prime}H')' + (F + GK_0^{*\prime}H')M + I = 0$$

$$K_0^{*\prime} = -R^{-1}G'\bar{P}MH(H'MH)^{-1}. \tag{13.2-15}$$

Several points can now be noted. First, for a given triple F, G, H, there may be no K_0 such that $F + GK_0'H'$ has all eigenvalues with negative real parts. (See Problem 13.2-4.) Hence, there may be no optimal control law K_0^*. Second, the unknown matrices in (13.2-15)—viz., K_0^*, $\bar{P}$, and M—are related in a highly nonlinear fashion; it is *not* possible, in general, to deduce from (13.2-15) a quadratic matrix equation for $\bar{P}$ or M whose solution then allows simple computation of the other two matrices. In fact, there is no technique for ready solution of (13.2-15). Third, precisely because of the nonlinearity of (13.2-15), it is not known whether there exist other K_0 than the optimal K_0^*, which, with an associated $\bar{P}$ and M, satisfy the equations. Consequently, any iterative technique leading to a solution of (13.2-15) might possibly not yield the truly optimal K_0^*. Fourth, in the special case when H^{-1} exists, (13.2-15) reduces to the standard regulator problem (see Problem 13.2-5). This is, of course, as one might expect, since with H^{-1} existing, the states can be recovered from the outputs.

We shall now briefly consider *procedures for computing K_0^*, $\bar{P}$, and M.* The two following procedures may lead to spurious solutions of (13.2-15)—i.e., solutions that do not define the optimal K_0^*; furthermore, neither procedure is actually guaranteed to converge. The first procedure, [3], is as follows. (Brief remarks follow each step of the procedure.)

1. Choose $K_0^{(0)}$ such that $F + GK_0^{(0)\prime}H'$ has all eigenvalues with negative real part. [The construction of such a $K_0^{(0)}$ appears to be a difficult problem in general, for which no explicit procedure is known. Of course, if F has all eigenvalues with negative real parts, $K_0^{(0)} = 0$ is satisfactory.]
2. Define $\bar{P}^{(i)}$ from $K_0^{(i-1)}$ as the solution of
$$\bar{P}^{(i)}[F + GK_0^{(i-1)\prime}H'] + [F + GK_0^{(i-1)\prime}H']'\bar{P}^{(i)} + HK_0^{(i-1)}RK_0^{(i-1)\prime}H' + Q = 0.$$
(This equation is linear and comparatively easy to solve.)
3. Define $M^{(i)}$ and $K_0^{(i)}$ from $\bar{P}^{(i)}$ by simultaneous solution of
$$M^{(i)}[F + GK_0^{(i)\prime}H']' + [F + GK_0^{(i)\prime}H']M^{(i)} + I = 0 \tag{13.2-16}$$
$$K_0^{(i)\prime} = -R^{-1}G'\bar{P}^{(i)}M^{(i)}H[H'M^{(i)}H]^{-1}. \tag{13.2-17}$$
More precisely, substitute for $K_0^{(i)}$ in (13.2-16), using (13.2-17). This yields a nonlinear equation for $M^{(i)}$; solve this equation, and then obtain $K_0^{(i)}$ from (13.2-16). (There seems to be no quick technique for tackling the nonlinear equation.)
4. Check that $[F + GK_0^{(i)\prime}H']$ has all eigenvalues with negative real parts. If this is the case, return to (2). (If not, the algorithm fails.)
5. Take $K_0^* = \lim_{i\to\infty} K_0^{(i)}$, etc.

Although this procedure is not guaranteed to converge, either to the optimal

or to any solution of (13.2-15), it does guarantee that $\operatorname{tr}[\bar{P}^{(i)}]$ converges to a limit as $i \longrightarrow \infty$ (see [3]). Of course, the fact that $\operatorname{tr}[\bar{P}^{(i)}]$ converges does not imply that the $\bar{P}^{(i)}$ converge, as the following example shows:

$$\bar{P}^{(i)} = \begin{bmatrix} 2 + (-1)^n & 0 \\ 0 & 1 + \dfrac{1}{n^2} - (-1)^n \end{bmatrix}.$$

However, one suspects that convergence of $\operatorname{tr}[\bar{P}^{(i)}]$ in some way will "encourage" convergence of $\bar{P}^{(i)}$.

The second computational scheme removes the difficult step of solving a nonlinear equation to obtain $M^{(i)}$, which is required in the first computational scheme. Only linear equations need be solved. However, the penalty paid is that no information regarding convergence is known—not even concerning $\operatorname{tr}[\bar{P}^{(i)}]$. As for the first scheme, the algorithm breaks down if at each stage a certain matrix fails to have all its eigenvalues with negative real parts. An outline of the scheme is as follows:

1. Choose $K_0^{(0)}$ such that $F + GK_0^{(0)\prime}H'$ has all eigenvalues with negative real parts.
2. Define $\bar{P}^{(i)}$ and $M^{(i)}$ from $K_0^{(i-1)}$ as the solutions of

$$\bar{P}^{(i)}[F + GK_0^{(i-1)\prime}H'] + [F + GK_0^{(i-1)\prime}H']'\bar{P}^{(i)} + HK_0^{(i-1)}RK_0^{(i-1)\prime}H' + Q = 0$$

and

$$M^{(i)}[F + GK^{(i-1)\prime}H']' + [F + GK^{(i-1)\prime}H']M^{(i)} + I = 0.$$

3. Define $K^{(i)}$ from $\bar{P}^{(i)}$ and $M^{(i)}$ as

$$K_0^{(i)\prime} = -R^{-1}G'\bar{P}^{(i)}M^{(i)}H[H'M^{(i)}H]^{-1}.$$

4. Check that $F + GK_0^{(i)\prime}H'$ has all eigenvalues with negative real parts. If this is the case, return to (2). If not, the algorithm fails.
5. Take $K_0^* = \lim_{i \to \infty} K_0^{(i)}$, etc.

Other approaches to computation could perhaps be based on gradient procedures, with $\partial[\operatorname{tr} \bar{P}]/\partial K_0$ even being computed experimentally by perturbing K_0.

An example drawn from [3] is the following. The system given has

$$F = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad g = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad h = [0 \quad 1]$$

and is evidently stabilizable by memoryless output feedback. The Q matrix is

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

whereas $R = [1]$. This leads to the optimal control law $u = -\sqrt{\frac{2}{3}}\,y$. That is, $K^* = -\sqrt{\frac{2}{3}}$, with

$$\bar{P} = \begin{bmatrix} 1.4 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

and the optimal performance index $= \operatorname{tr}[\bar{P}] = 2.4$. With state feedback, the control $u = [-0.4 \quad -0.9]x$ is optimal, and the optimum performance index is

$$x'(0)\begin{bmatrix} 1.3 & 0.4 \\ 0.4 & 0.9 \end{bmatrix}x(0).$$

Averaging over all points on the unit sphere, we obtain 2.2 for the averaged performance index.

Problem 13.2-1. Explain how the situation where $H'\Phi\Phi'H$ is singular can be dealt with, to deduce an equation akin to (13.2-14). [This also applies to (13.2-15).]

Problem 13.2-2. Show that the conditions (13.2-12) are sufficient for optimality when the controller gain may be time varying. (*Hint:* Use the Hamilton–Jacobi theory of Chapter 2. Let $X(t)$ be an $n \times n$ matrix satisfying

$$\dot{X} = (F + GK')X \qquad X(t_0) \quad \text{given.}$$

We regard the entries of X, arranged as a vector if desired, as the entries of the state vector, and the entries of K as the entries of a control vector. As the performance index $V(X(t), K(\cdot), t)$ (the time variable referring now to the initial time), we take

$$V(X(t), K(\cdot), t) = \operatorname{tr} \int_t^T X'(\tau)[K(\tau)R(\tau)K'(\tau) + Q(\tau)]X(\tau)\,d\tau,$$

which we hypothesize has minimum value

$$V^*(X(t), t) = \operatorname{tr}[X'(t)P(t)X(t)]$$

for some nonnegative definite matrix $P(t)$. Verify that $\partial V^*/\partial X(t)$, which is a matrix whose i-j entry is $\partial V/\partial X_{ij}(t)$, is $2P(t)X(t)$. The Hamilton–Jacobi equation is

$$\begin{aligned}\frac{\partial}{\partial t} V^*(X(t), t) = -\min_{K_0(t)} \{&\operatorname{tr}[X'(HK_0RK_0'H' + Q)X] \\ &+ \sum_{i,j} (2PX)_{ij}(FX + GK_0'H'X)_{ij}\}\end{aligned}$$

where the argument of the matrices within $\{\cdots\}$ has been omitted for clarity. Check that this may be rewritten

$$\frac{\partial}{\partial t} V^*(X(t), t) = -\min_{K_0(t)} \{\operatorname{tr}[X'(HK_0RK_0'H' + Q + 2PF + 2PGK_0'H')X]\}.$$

Verify also that

$$\begin{aligned}\frac{\partial}{\partial K_0} &\operatorname{tr}[X'(HK_0RK_0'H' + Q + 2PF + 2PGK_0'H')X] \\ &= 2(RK_0'H'XX'H + G'XX'H).\end{aligned}$$

Setting this derivative equal to zero allows computation of the minimizing K_0. Verify that when this minimizing K_0 is inserted into the Hamilton–Jacobi equation, and the left side of the equation is evaluated, then the differential equation given in the text for $P(t)$ is recovered.)

Problem 13.2-3. Derive directly the necessary conditions for optimality (13.2-15) without specializing the finite-time results derived earlier in the section. How would the conditions be changed if the closed-loop system were constrained to have a prescribed degree of stability of at least α? What happens when the index is $(1/n)\,\mathrm{tr}[A\bar{P}]$ for some positive definite symmetric A?

Problem 13.2-4. Show that there exist matrix triples F, G, H, such that $F + GK_0'H'$ can never have all eigenvalues with negative real part, irrespective of the choice of K_0. For ease, consider a single-input, single-output system. (Reference [5] gives such an example, which has the interesting property that a time-varying gain K can be found such that the closed-loop system is stable.)

Problem 13.2-5. Show that the assumption that H is invertible in the equation set (13.2-15) leads to these equations being essentially the standard regulator equations.

13.3 PROPERTIES OF OPTIMAL SOLUTIONS

Throughout this book, we have stressed the fact that optimality per se may not be important, but ancillary properties associated with certain optimal solutions may be important. We recall that for the standard optimal regulator there are a number of important and pleasing properties associated with optimality, and so we are led to ask what are the ancillary advantages of using feedback, particularly output feedback, which is optimal in the sense of this chapter. First and obviously, the controller constraints we have incorporated into the designs normally lead to controllers that are simpler to implement physically than a controller design using state-variable feedback, as in a standard optimal regulator. Second, the closed-loop system resulting from optimal output feedback is asymptotically stable—at least for the infinite-time case considered. This is hardly a significant advantage, however; in contrast to the state-variable feedback case, one cannot use the optimal design procedure as a technique for stabilizing an unstable open-loop system, since the first step in the computation of the optimal output feedback law generally requires the choice of some output feedback law stabilizing the system.

It appears that it may not be possible to make general statements about the properties of the nonstandard optimal systems, in which case it would be helpful to have at least some indication of how good the closed-loop systems

really are. It might be argued that the index $\mathrm{tr}[\bar{P}]$ gives us such a measure, particularly if we have available the value of $\mathrm{tr}[P^*]$ associated with the corresponding standard optimal regulator. (The closer the two indices, the closer we might expect the system properties to be to one another.)

For the case when an infinite-time optimization has been carried out, it is possible to do a little better than the preceding comparison by considering the extent of tolerance of nonlinearities in the plant input transducers and relating this to what is already known for the standard regulator. This eliminates the need to solve any corresponding standard regulator problem.

Tolerance of nonlinearities in a closed-loop system. We recall that arbitrary nonlinearities in the sector $[\frac{1}{2}, \infty]$ can be accommodated in the standard optimal regulator without loss of the guaranteed asymptotic stability properties. Associated with this property are good sensitivity properties, good phase margin, tolerance of time delays, etc. We now show that for any nonstandard optimal system, it is possible to calculate a sector $[\kappa_1, \kappa_2]$, normally with $\kappa_1 > \frac{1}{2}$, $\kappa_2 < \infty$, such that nonlinearities in this sector do not disturb system stability. The interval $[\kappa_1, \kappa_2]$ is then a performance measure that can be directly interpreted in physical terms. Moreover, the knowledge of $[\kappa_1, \kappa_2]$ gives us an indication of how "close" a particular nonstandard regulator system is to the standard one—without having to solve the corresponding standard regulator problem. Actually, the results that will be developed are applicable, with modification, to any system which is not a standard optimal one, but they are presented here for the case when we have a nonstandard optimal system design and we desire to evaluate its merits.

Let us consider the time-invariant plant

$$\dot{x} = Fx + Gu \qquad y = H'x$$

with a *nominal* control law $u = K'x$. This control law has the form $K = HK_0$ for some K_0 and is optimal in the sense described in the previous section. That is, the performance index $\mathrm{tr}[\bar{P}]$ is minimized where $\bar{P}$ satisfies the linear equation

$$\bar{P}(F + GK') + (F + GK')'\bar{P} + (KRK' + Q) = 0$$

for some positive definite symmetric R and nonnegative definite symmetric Q. Implicit in what we have assumed is that the closed-loop system

$$\dot{x} = (F + GK')x$$

is asymptotically stable. For ease in subsequent calculation, we assume that $KRK' + Q$ is nonsingular, thus guaranteeing that $\bar{P}$ is positive definite.

Let us suppose that, in fact, the feedback law is $u = \phi(K'x, t)$ where the vector $\phi(\cdot, \cdot)$ represents nonlinearities (or even time-varying gains) in the system input transducers. We now investigate the stability properties of

the closed-loop system

$$\dot{x} = Fx + G\phi(K'x, t). \tag{13.3-1}$$

In particular, we seek sector bounds $[\kappa_1, \kappa_2]$, such that if

$$(\kappa_1 + \epsilon)y'y \leq \phi'(y, t)y \leq (\kappa_2 - \epsilon)y'y \tag{13.3-2}$$

for all y and t and some positive ϵ, the closed-loop system (13.3-1) is asymptotically stable.

For the general multiple-input problem, it appears that explicit formulas for κ_1 and κ_2 are not available, although κ_1 and κ_2 can be defined implicitly. In an effort to obtain explicit results, we shall therefore restrict ourselves somewhat. One obvious restriction to impose is to constrain $\phi(\cdot, \cdot)$ to be a diagonal matrix, or, equivalently, to decouple any nonlinearities in the input transducers. This assumption is reasonable, but unfortunately still leads to some difficulty in any attempt to calculate explicit expressions for the bounds $[\kappa_1, \kappa_2]$. If, however, we now make the further restriction that $\phi(K'x, t)$ has the form $\kappa(t)K'x$ for some scalar gain $\kappa(t)$, *whose time-variation is perhaps due to nonlinearities*, then the bounds $[\kappa_1, \kappa_2]$—now bounds on $\kappa(t)$—may be found explicitly. This restriction is admittedly unrelated to any usual practical situation, except, of course, for single-input systems,† when it is no restriction at all. However, it does enable a quick computation to be made for $[\kappa_1, \kappa_2]$.

With the assumption that $\phi(K'x, t)$ is of the form $\kappa(t)K'x$, some scalar $\kappa(\cdot)$, we now show that the bounds $[\kappa_1, \kappa_2]$ are given explicitly by

$$\kappa_1 = 1 + (\lambda_{\min}[E])^{-1} \tag{13.3-3}$$

$$\kappa_2 = 1 + (\lambda_{\max}[E])^{-1} \tag{13.3-4}$$

where

$$E = [\bar{P}G \quad K]'(KRK' + Q)^{-1}[K \quad \bar{P}G]. \tag{13.3-5}$$

Although the expressions for the bounds κ_1 and κ_2 appear to be complicated, they are, in fact, not really difficult to calculate once $\bar{P}$ has been determined as part of the nonstandard optimal design. In fact, for the case when we have a single-input system (G replaced by g, K replaced by k, etc.), the eigenvalues $\lambda_{\min}[E]$ and $\lambda_{\max}[E]$ may be computed to be

$$\begin{aligned}\lambda_{\max}[E], \lambda_{\min}[E] = {} & k'(kk' + Q)^{-1}\bar{P}g \\ & \pm [g'\bar{P}(kk' + Q)^{-1}\bar{P}gk'(kk' + Q)^{-1}k]^{1/2}\end{aligned} \tag{13.3-6}$$

Problem 13.3-1 asks for verification of this result, given Eq. (13.3-5).

To investigate the stability properties of the closed-loop system (13.3-1),

†For single-input systems, sector bounds $[\kappa_1, \kappa_2]$ can also be determined by using a Nyquist plot together with the Circle Criterion.

we adopt as a tentative Lyapunov function

$$V = x'\bar{P}x.$$

Since $\bar{P}$ is positive definite, it remains for us to examine $\dot{V}$.

Differentiating yields

$$\begin{aligned}\dot{V} &= x'\bar{P}\dot{x} + \dot{x}'\bar{P}x \\ &= x'[\bar{P}(F + \kappa(t)GK') + (F + \kappa(t)GK')'\bar{P}]x \\ &= -x'(KRK' + Q)x + (\kappa(t) - 1)x'[\bar{P}GK' + KG'\bar{P}]x.\end{aligned}$$

The function $\dot{V}$ is negative definite for all time, provided that the inequality

$$(KRK' + Q) - (\kappa - 1)[\bar{P}GK' + KG'\bar{P}] \geq \epsilon I > 0 \tag{13.3-7}$$

holds for all time and some positive constant ϵ. (The notation $\geq$ [$>$] is shorthand for nonnegative [positive] definite.) The inequality (13.3-7) thus guarantees asymptotic stability of the closed-loop system (13.3-1).

We now introduce the identifications

$$A = (KRK' + Q)^{-1/2}\bar{P}G \qquad B = (KRK' + Q)^{-1/2}K.$$

These imply that

$$E = \begin{bmatrix} A'B & A'A \\ B'B & B'A \end{bmatrix}$$

where E is as defined earlier. The preceding identifications enable us to write the inequality (13.3-7) as

$$I - (\kappa - 1)[AB' + BA'] \geq \epsilon I > 0$$

which is equivalent to requiring that the nonzero eigenvalues of $[AB' + BA']$, written $\lambda_i[AB' + BA']$, satisfy

$$(\kappa - 1)\lambda_i[AB' + BA'] \leq (1 - \epsilon)I$$

for all i. But it may readily be shown that the nonzero eigenvalues of $[AB' + BA']$ are identical to the nonzero eigenvalues of E. (See Problem 13.3-2.) The preceding inequality is thus equivalent to

$$(\kappa - 1)\lambda_i[E] \leq (1 - \epsilon)I \tag{13.3-8}$$

for all i. For the usual case, when $\lambda_{\min}[E]$ and $\lambda_{\max}[E]$ are opposite in sign, it is immediate that this inequality is equivalent to

$$(1 - \epsilon)(\lambda_{\min}[E])^{-1} \leq \kappa - 1 \leq (1 - \epsilon)(\lambda_{\max}[E])^{-1}.$$

The bounds on $\kappa(t)$ given by (13.3-3) and (13.3-4) are now immediate. In case $\lambda_{\min}[E]$ is not negative ($\lambda_{\max}[E]$ is not positive), then the bound κ_1 [κ_2] ceases to exist.

Problem 13.3-3 requests the reader to show that when $K = -\bar{P}G$, corresponding to the standard optimal regulator case, sector bounds of $\frac{1}{2}$ and ∞ result from these calculations.

To conclude this section, we give some indication of the results possible in the general multiple-input case. Suppose $K'x$ is an m vector. We consider a class of $\phi(\cdot, \cdot)$ defined through an $m \times m$ matrix $\Phi(t)$:

$$\phi(K'x, t) = \Phi(t)K'x. \tag{13.3-9}$$

[The time variation in $\Phi(t)$ may be due to nonlinearities rather than to variation of some parameter. However, the argument $K'x$ is suppressed in $\Phi(t)$.]

As before, a tentative Lyapunov function $V = x'\bar{P}x$ is adopted, and $\dot{V}$ is evaluated to be

$$-x'[KRK' + Q + \bar{P}G(\Phi - 1)K' + K(\Phi' - 1)G'\bar{P}]x.$$

Asymptotic stability follows if $\dot{V}$ is negative definite, which follows if

$$I - [A(\Phi - I)B' + B(\Phi' - I)A'] \geq \epsilon I > 0$$

in terms of the earlier definitions for A and B. In turn, this inequality holds if

$$\lambda_{\max}\left[\begin{bmatrix} \Phi - I & 0 \\ 0 & \Phi' - I \end{bmatrix} E\right] \leq 1 - \epsilon. \tag{13.3-10}$$

Rewriting of this inequality in terms of specific bounds on Φ seems very difficult. However, the inequality may be of some benefit practically.

Examples have shown for the scalar-input case that the bounds of (13.3-3) and (13.3-4) are highly conservative; less conservative bounds are derivable by use of a Nyquist plot in conjunction with the Circle Criterion. For the multiple-input case however, the above theory, though incomplete, appears to represent the only feasible approach. Presentation of the theory appears justified on these grounds alone.

Problem 13.3-1. For the single-input case, verify that the result (13.3-6) follows from (13.3-5). Show also that $\lambda_{\min}[E]$ and $\lambda_{\max}[E]$ can never have the same sign.

Problem 13.2-2. Show that the nonzero eigenvalues of $AB' + BA'$ are the same as those of

$$\begin{bmatrix} A'B & A'A \\ B'B & B'A \end{bmatrix}.$$

Problem 13.3-3. Using the notation of the section, show that when

$$K = -\bar{P}G$$

(the standard optimal regulator case) and $Q > 0$, the sector bounds $[\kappa_1, \kappa_2]$ satisfy $\kappa_1 \leq \frac{1}{2}$, $\kappa_2 = \infty$. [*Hint:* First show that the inequality $K'[Q + KK']^{-1}K \leq I$ must be satisfied. Show also that $K'(\alpha I + KK')^{-1}K = K'K(\alpha I + K'K)^{-1}$, and use the fact that $(Q + KK') \geq (\alpha I + KK')$ for some α.]

Problem 13.3-4. Suppose $\dot{x} = Fx + gu$ is a prescribed single-input, time-invariant system, and $u = k'x$ is a control chosen to stabilize the system; note

that $u = k'x$ is not assumed to result from an optimal design procedure. Show how sector bounds for nonlinearity tolerance can be calculated in terms of an arbitrary, positive definite Q, and a matrix P satisfying

$$P(F + gk') + (F + gk')'P + kk' + Q = 0.$$

The sector bounds depend on the choice of Q. Can you suggest how Q might be chosen?

13.4 DYNAMIC CONTROLLERS

In this section we consider, somewhat briefly, the optimal design of dynamic controllers for linear regulators where constraints are imposed on the order of the controller. As may be inferred from the theory of earlier chapters, two possible approaches may be used. In the first approach, some of the system states are estimated with a dynamic estimator; these state estimates are then used in a nondynamic state feedback control law. The state estimator and nondynamic control law together constitute the optimal dynamic controller for the original system, and both must be optimally designed. In the second approach, a performance index is taken, which includes derivatives of the control variable; the plant is turned into an augmented system by the insertion of integrators at the input, and the optimal feedback law for the augmented system is derived. Constraints requiring output feedback will carry over from the original system to the augmented system, but the augmented system feedback law can also include feedback of the original system input (which becomes an augmented system state). The theory of earlier sections of this chapter, of course, applies to the selection of the augmented system feedback law, which can then be translated into a dynamic law for the original system.

Design using estimators: the estimation problem. We shall consider the case of an n-dimensional, time-invariant, single-input, single-output plant

$$\dot{x} = Fx + gu \quad y = h'x. \tag{13.4-1}$$

We also consider a state estimator having state equations

$$\dot{w} = F_e w + g_{1e}u + g_{2e}y \tag{13.4-2}$$

where the dimension of F_e is less than or equal to the dimension of F. If the dimension of F_e is the same as that of F, estimators that are optimal in the presence of noise may be designed; if F_e has dimension one less than that of F, certainly a full estimate of x is achievable. Here, we are more interested in the case when the dimension of F_e is less than that of F by two or more (perhaps because of a design constraint on the complexity of the controller). In any case, the vector w is required to approach asymptotically a vector

of the form $H_1'x$, for a certain H_1. That is, we require the response of the system (13.4-1) and estimator (13.4-2) to approximate the response of the system

$$\dot{x} = Fx + gu \qquad y_a = [h \quad H_1]'x \tag{13.4-3}$$

for some matrix H_1.

For w to be an estimate of H'_1x, it is clear, from rewriting (13.4-1) and (13.4-2) as

$$(\dot{w} - H'_1\dot{x}) = F_e(w - H'_1x) + (g_{1e} - H'_1g)u$$
$$+ (g_{2e}h' - H'_1F + F_eH'_1)x,$$

that F_e must be chosen such that the system $\dot{z} = F_ez$ is asymptotically stable, and g_{1e}, g_{2e}, F_e, and H_1 must satisfy

$$(g_{1e} - H'_1g) = 0 \qquad (g_{2e}h' - H'_1F + F_eH'_1) = 0. \tag{13.4-4}$$

If F, h, F_e, and g_{2e} are specified, the second of Eqs. (13.4-4) becomes a linear matrix equation for the unknown H_1. It is always solvable, as described in Appendix A, if F and F_e have no eigenvalues in common. Once H_1 is determined, g_{1e} follows immediately from the first of Eqs. (13.4-4).

The question remains as to how F_e and g_{2e} should be chosen. Some immediate restrictions on these matrices are that F_e should have eigenvalues sufficiently far in the left half-plane to ensure fast estimation, but not so far in the left half-plane that noise becomes a problem; furthermore, a choice of F_e and g_{2e} should be rejected if it leads to scalar outputs present in H'_1x not being linearly independent of the output $h'x$. (Otherwise, the estimator is needlessly complex—unless, of course, there is some advantage associated with filtering of contaminating noise.) But in the final analysis, the selection of F_e and g_{2e} is determined by the resulting H_1, which, in turn, is linked to the control problem that we consider now.

Design using estimators: the control problem. Let us suppose that the estimator dimension is fixed and, for the moment, that the estimator itself is fixed. We consider temporarily the system (13.4-3); with the aid of the theory of previous sections, an optimal control law of the following variety can be formed:

$$u^* = [k'_1 \quad k'_2]y_a.$$

Now, the control law $u = k'_1y + k'_2w$ is certainly a good candidate to consider for controlling the original system (13.4-1) augmented with the estimator (13.4-2). The dynamic controller for this case has the state equations

$$\dot{w} = (F_e + g_{1e}k'_2)w + (g_{2e} + g_{1e}k'_1)y \tag{13.4-5}$$

$$u = k'_1y + k'_2w. \tag{13.4-6}$$

We expect that this controller can achieve a response that becomes closer to that of a controller consisting of a standard Luenberger observer together with standard state-variable feedback as the rank of $[h \quad H_1]$ approaches n. For the particular case when rank $[h \quad H_1]$ is n, the preceding dynamic controller has the form of the Luenberger estimator, which estimates H'_1x, followed by a constant control law, which is readily derived from the standard optimal state feedback law as follows. If $u^* = k'x$ is the standard optimal law, then the gains $[k'_1 \quad k'_2]$ are given from

$$[k'_1 \quad k'_2] = k'\begin{bmatrix} h' \\ H'_1 \end{bmatrix}^{-1}.$$

The derivation of the estimator parameters such that rank $[h \quad H_1] = n$ is, of course, the difficult part of standard estimator design, as we have seen in Chapter 8 where the problem is approached from an alternative viewpoint.

We now suppose that rank $[h \quad H_1]$ is not necessarily n, and that the estimator is no longer fixed. Our theory so far has avoided the important question as to what are the best values that can be chosen for g_{2e} and F_e. Since different values of these matrices lead to different values of H_1, and, in turn, to different optimal control laws and performance indices, the real problem becomes one of choosing g_{2e} and F_e so that the performance index is minimized [*Note:* The performance index measures the performance of (13.4-3), *not* (13.4-1), or (13.4-1) and (13.4-2) together. We obtain a control law that exactly minimizes this index, and then construct from it a dynamic control law for (13.4-1), which is not actually optimal for (13.4-1) but obviously has the right sort of properties.] *It turns out that we can extend the algorithm used for calculating the optimum k_1 and k_2, knowing H_1, to incorporate the determination of k_1, k_2, g_{2e}, and F_e within the required constraints.* Problem 13.4-1 asks that the case of a single-input, single-output plant with first-order compensator (i.e., w is a scalar) be studied in detail.

Since the writing of this section, an approach has been developed to the design of controllers which uses a performance index involving both the controller parameters and states [7]. The advantage of the approach to the design of dynamic controllers just given when compared with the approach of [7], and for that matter the approach next described, is that the quadratic form $(u'Ru + x'Qx)$ occurring in the performance index remains unchanged as different controllers, including controllers of different dimension, are considered.

Design using control derivatives in a performance index. In the latter part of this section on dynamic controllers, we consider the approach that combines some of the ideas of Chapter 10 with the results of earlier sections of this chapter. This particular approach consists in optimally controlling the original system when augmented with dynamics at the inputs, using a

nondynamic control law. The optimal augmented system with a nondynamic controller is then interpreted as the original system with an optimal dynamic controller. The index minimized for the augmented system, when interpreted in terms of an index for the original system, includes cost terms involving derivatives of the original system inputs.

One minor difficulty with this approach is that as the order of the compensator or controller changes, the form of the performance index associated with the original system must change, and thus there is no way of comparing two controllers of different order. This is not the case for the estimator approach discussed in the earlier part of this section.

We now illustrate the linkage of the ideas of Chapter 10 and the results of earlier sections of this chapter with an example taken from [8]. (Incidentally, the problem considered in the example cannot be solved using the estimator approach. The system considered has a fixed dynamic structure that cannot be viewed as an estimator in series with a nondynamic control law.)

EXAMPLE. *System design using a fixed structure controller.* Suppose we are given a single-input, single-output plant

$$\dot{x} = Fx + gu \quad y = h'x. \tag{13.4-7}$$

Let us suppose also that we wish to control the system using a fixed structure controller having equations

$$u = k_1 \int_0^t y\,dt + k_2 y + k_3 \dot{y} \tag{13.4-8}$$

where the constants k_1, k_2, and k_3 are adjustable parameters. This controller structure is often referred to as a *three-term controller* for obvious reasons, and has found wide application for stabilizing industrial control systems; in practice, the derivative in (13.4-8) is obtained approximately.

We desire to select the parameters $k = [k_1, k_2, k_3]$ so that the closed-loop system stability properties are as close as possible to those of a standard regulator. As a first step in formulating the problem in optimal control terms, we express the control law (13.4-8) as a state feedback law using (13.4-7). We have from (13.4-7) that

$$y = h'x \qquad \dot{y} = h'Fx + h'gu \qquad \ddot{y} = h'F^2x + h'Fgu + h'g\dot{u}$$

which means that (13.4-8), when differentiated, may be written

$$(1 - k_3h'g)\dot{u} - (k_3h'F^2 + k_2h'F + k_1h')x - (k_3h'Fg + k_2h'g)u = 0.$$

Using the notation $\hat{k}$ as a normalization of k as follows,

$$\hat{k} = \begin{bmatrix} \hat{k}_1 \\ \hat{k}_2 \\ \hat{k}_3 \end{bmatrix} = (1 - k_3h'g)^{-1} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \tag{13.4-9}$$

we may write the preceding control law as

$$\dot{u} = \hat{k}'[h \quad F'h \quad (F')^2 \quad h]'x + \hat{k}'[0 \quad g'h \quad g'F'h]'u. \tag{13.4-10}$$

We now recall, using the ideas of Chapter 10, that this is nothing other than a state feedback law for the system (13.4-7), augmented with an integrator at its input. That is, (13.4-10) may be written as

$$u_a = k_a' x_a \tag{13.4-11}$$

where

$$u_a = \dot{u} \qquad x_a = \begin{bmatrix} x \\ u \end{bmatrix} \qquad k_a = \begin{bmatrix} h & F'h & (F')^2 h \\ 0 & g'h & g'F'h \end{bmatrix} \hat{k}. \tag{13.4-12}$$

The augmented system equations are given from (13.4-7) and (13.4-12) as

$$\dot{x}_a = F_a x_a + g_a u_a$$

where

$$F_a = \begin{bmatrix} F & g \\ 0 & 0 \end{bmatrix} \qquad g_a = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{13.4-13}$$

So far, all we have done is show how the three-term controller can be viewed as a state-variable feedback law for the original system augmented with an integrator at its input. If the gain elements k_1, k_2, k_3 of a three-term controller are specified, then $\hat{k}$ may be calculated by using (13.4-9); also, the state feedback gain k_a for the augmented system (13.4-13) may be calculated by using (13.4-12).

Now we are in a position to apply the results of this chapter to the augmented system. A performance index for the augmented system may be defined and then k_1, k_2, and k_3 adjusted until this index is minimized.

We now give further details for the case when the index is

$$\bar{V} = \operatorname{tr}[\bar{P}]$$

where

$$\bar{P} = \int_0^\infty \exp[(F_a + g_a k_a')' t](Q + k_a k_a') \exp[(F_a + g_a k_a')t]\, dt$$

for some positive definite symmetric Q. For this case, certainly $\partial \bar{V}/\partial k_a$ can be calculated using results of earlier sections. In fact, it is not hard to calculate $\partial \bar{V}/\partial k$ rather than $\partial \bar{V}/\partial k_a$, using minor extensions of the earlier results. Setting this to zero gives necessary conditions for optimality. (See Problem 13.4-2.)

For the case when

$$F = \begin{bmatrix} 0 & 1 & 10 \\ 0 & 0 & 1 \\ 0 & -9 & -10 \end{bmatrix} \quad g = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad h' = [1 \quad 0 \quad 0] \quad Q = I$$

the optimal feedback law has been calculated to be

$$k_1 = 0.2 \quad k_2 = 2.21 \quad k_3 = 1.38.$$

The minimum index is $\bar{V}^* = 7.64$. This may be compared to that for the standard optimum (augmented) regulator, where $\bar{V}^* = 6.31$. Using the results of

the previous section, we calculate the sector in which nonlinearities may be tolerated in the plant input transducers as $[\kappa_1, \kappa_2]$, where

$$\kappa_1 = 0.7 \quad \kappa_2 = 1.3.$$

(Note that the theory of Chapter 10 has been assumed, which notes that nonlinearities that can be accommodated at the augmented plant input can also be accommodated at the original plant input.) These bounds are actually very conservative, as may be checked.

We comment that it is somewhat surprising that so much theory is necessary to select "optimally" the parameters of a three-term controller—particularly since in practice three-term controller parameters are selected for quite complex systems with often very little effort and quite good results [9].

In conclusion, we emphasize that many of the results discussed throughout the chapter are recent, and, as we have seen, there is still considerable scope for further developments. In fact since the writing of this section, a simpler result has been developed (in collaboration with D. Williamson and T. Fortmann) which is of considerable assistance in determining minimal dimension controllers—see Problem 13.4-3.

Problem 13.4-1. Consider the case of a single-input, single-output system

$$\dot{x} = Fx + gu \quad y = h'x$$

with a first-order estimator at its output, having state equations

$$\dot{w} = f_e w + g_{1e} u + g_{2e} y$$

where

$$g_{1e} = h_1' g \qquad g_{2e} h' - h_1' F + f_e h_1' = 0$$

and $\dot{z} = f_e z$ is asymptotically stable. Associated with this system is the following system

$$\dot{x} = Fx + gu \qquad y_a = [h \quad h_1(g_{2e}, f_e)]' x$$

where, according to the theory of this section, the estimator states w approach $h_1' x$ asymptotically. Define a suitable performance index of the form $\bar{V} = \operatorname{tr}[\bar{P}]$ for this system, and evaluate $\partial \bar{V}/\partial f_e$. Indicate how this result can be used in determining an optimal dynamic controller.

Problem 13.4-2. For the three-term controller problem, derive the necessary conditions for optimality. That is, find an expression for $\partial \bar{V}/\partial k_a$ and set this equal to zero.

Problem 13.4-3. Consider the nth order system (13.4-1) and pth order controller (13.4-5 and 13.4-6) where, using the notation of Chapter 8, $\det(sI - F_e) = s^p + \alpha_p s^{p-1} + \ldots + \alpha_1$ and $T_1 F T_1^{-1}$ is in transpose companion matrix form for T_1 given in (8.3-10). With $[F_e, k_2]$ completely observable and $\dot{z} = F_e z$ asymptotically

stable show that a necessary and sufficient condition for u to be an estimate of $k'x$ for some fixed k is that $S(T_1^{-1})'k = 0$ with

$$S = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_p & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \alpha_1 & \cdots & \alpha_{p-2} & \alpha_{p-1} & \alpha_p & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \cdot & \cdot & \cdot & \cdots & \alpha_p & 1 & 0 \end{bmatrix}$$

[*Hint:* First show that a sufficient condition is that $k = hk_1 + H_1k_2$ for some H_1 which satisfies (13.4-4). Next show that $S[H_1 \quad h] = 0$, by using the Cayley-Hamilton theorem and the fact that (13.4-4) constrains H_1 to have the form $H_1 = [h_1 \quad F_eh_1 \ldots F_e^{n-1}h_1]$. Assuming that h_1 is chosen so that $[F_e, h_1]$ is completely controllable, conclude the result. The multiple output case results in similar conditions].

REFERENCES

[1] Kleinman, D. L., T. Fortmann, and M. Athans, "On the Design of Linear Systems with Piecewise-Constant Feedback Gains," *IEEE Trans. Auto. Control*, Vol. AC–13, No. 4, August 1965, pp. 354–361.

[2] Levine, W. S., and M. Athans, "Average-Optimal Output-Feedback Controllers for Linear Time-Varying Systems" (to be published).

[3] Levine, W. S., and M. Athans, "On the Determination of the Optimal Constant Output-Feedback Gains for Linear Multivariable Systems," *IEEE Trans. Auto. Control*, Vol. AC–15, No. 1, February 1970, pp. 44–48.

[4] Zoutendijk, G., "Nonlinear Programming, A Numerical Survey," *SIAM J. Control*, Vol. 4, No. 1, 1966, pp. 194–210.

[5] Brockett, R. W., and H. B. Lee, "Frequency-Domain Instability Criteria for Time-Varying and Nonlinear Systems," *Proc. IEEE*, Vol. 5, No. 5, May 1967, pp. 604–619.

[6] Kalman, R. E., P. L. Falb, and M. A. Arbib, *Topics in Mathematical System Theory*, McGraw-Hill Book Company, New York, 1969.

[7] Johnson, T. L., and M. Athans, "On the design of optimal constrained dynamic compensators for linear constant systems," (to be published).

[8] Williamson, D., and J. B. Moore, "Three-Term Controller Parameter Selection Using Suboptimal Regulator Theory" *IEEE Trans. Auto. Control*, Vol. AC–16, No. 1, February 1971.

[9] Koppel, L. B., *Introduction to Control Theory, with Application to Process Control*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968.

CHAPTER 14

TWO FURTHER TOPICS

14.1 DISCRETE TIME SYSTEMS

This section is devoted to a brief exposition of linear regulator design for discrete time systems. Our starting point is the state-space equation

$$x(t+1) = F(t)x(t) + G(t)u(t) \qquad x(t_0) \quad \text{given} \tag{14.1-1}$$

and the performance index

$$V(x(t_0), u(\cdot), t_0) = \sum_{t=t_0+1}^{t} [x'(t)Q(t)x(t) + u'(t-1)R(t)u(t-1)]. \tag{14.1-2}$$

In these equations, $x(t)$ is the state at time t and $u(t)$ the control at time t. Generally, but not always, t is assumed to take on integer values. The plant (14.1-1) is initially—i.e., at time t_0—in state $x(t_0)$, and the aim is to return the plant state to the origin, or a state close to the origin. To do this, we set up a performance index (14.1-2), in which $Q(t)$ and $R(t)$ are nonnegative definite symmetric matrices. [*Note:* We do *not* assume $R(t)$ to be positive definite.] The performance index has the property that "large" values of the state will tend to make the performance index large. Hence, by choosing the control sequence $u(t_0)$, $u(t_0 + 1), \ldots$, which minimizes the performance index, we can expect to achieve the desired regulator effect.

It might be thought curious that $R(t)$ is not positive definite, since in the corresponding continuous time performance index, the corresponding matrix is positive definite. In the latter case, the presence of the matrix rules out the possibility of using an infinitely large control to take the state to zero in an infinitely short time. In the discrete time case, it is not possible to take

the state to zero in an infinitely short time, and the possibility of an infinitely large control occurring does not even arise. Hence, there is no need to prevent such a possibility by using a positive definite $R(t)$.

We shall first solve the optimization problem for the case of finite T. Then, with additional assumptions, we shall cover the infinite T case, with special reference to time-invariant plants. Finally, we shall remark on such matters as stability, tracking, etc.

For finite T, we shall show that the optimal control is a state feedback law, and that the optimal performance index is quadratic in the initial state $x(t_0)$—results that are, of course, analogous with the corresponding continuous time results.

The route to a derivation of the optimal control is via the Principle of Optimality. Thus, if until time t optimal controls $u(t_0), u(t_0 + 1), \ldots, u(t - 1)$ have been applied, leading to a state $x(t)$, then the remaining terms in the optimal control sequence, $u(t), u(t + 1), \ldots, u(T - 1)$ must also be optimal in the sense of minimizing $V[x(t), u(\cdot), t]$.

Now let $V^*[x(t), t]$ denote the optimal performance index associated with an initial state $x(t)$ at time t. Then, by the Principle of Optimality,

$$\begin{aligned}
V^*[x(t), t] = \min_{u(t)} \{&[F(t)x(t) + G(t)u(t)]'Q(t + 1)[F(t)x(t) + G(t)u(t)] \\
&+ u'(t)R(t + 1)u(t) \\
&+ V^*(F(t)x(t) + G(t)u(t), t + 1)\} \\
= \min_{u(t)} \{&u'(t)[G'(t)Q(t + 1)G(t) + R(t + 1)]u(t) \\
&+ 2x'(t)F'(t)Q(t + 1)G(t)u(t) + x'(t)F'(t)Q(t + 1)F(t)x(t) \\
&+ V^*(F(t)x(t) + G(t)u(t), t + 1)]. \qquad (14.1\text{-}3)
\end{aligned}$$

Bearing in mind the corresponding continuous time results, *it would be reasonable to guess that $V^*[x(t), t]$ would be of the form $x'(t)P(t)x(t)$.* Since it proves convenient to make use of this result almost immediately, we build into the following argument an inductive proof of the result.

We require first the following assumption.

Assumption 14.1-1. For all t, $G'(t - 1)Q(t)G(t - 1) + R(t)$ is positive definite.

This assumption may, in fact, be relaxed. However, when it does not hold, the optimal control law becomes nonunique (although still linear), and we wish to avoid this complication. Notice that the assumption will very frequently hold, e.g., if $R(t)$ is positive definite for all t, or if $Q(t)$ is positive definite for all t and the columns of $G(t)$ are linearly independent, etc.

With Assumption 14.1-1, it is easy to evaluate the "starting point" for the induction hypothesis—viz., $V^*[x(T - 1), T - 1]$. We have

$$V(x(T - 1), u(\cdot), T - 1) = x'(T)Q(T)x(T) + u'(T - 1)R(T)u(T - 1)$$

and, in view of the system equation (14.1-1), this becomes

$$\begin{aligned}V(x(T-1), u(\cdot), T-1) = {} & x'(T-1)F'(T-1)Q(T)F(T-1)x(T-1) \\ & + 2x'(T-1)F'(T-1)Q(T)G(T-1)u(T-1) \\ & + u'(T-1)[G'(T-1)Q(T)G(T-1) \\ & + R(T)]u(T-1).\end{aligned}$$

Evidently, the control $u(T-1)$ that minimizes this performance index is a linear function of $x(T-1)$—i.e.,

$$u^*(T-1) = K'(T-1)x(T-1) \tag{14.1-4}$$

for a certain matrix $K(T-1)$. Moreover, the resulting optimal index $V^*(x(T-1), T-1)$ becomes quadratic in $x(T-1)$—i.e.,

$$V^*(x(T-1), T-1) = x'(T-1)P(T-1)x(T-1) \tag{14.1-5}$$

for a certain nonnegative definite symmetric $P(T-1)$. The actual expressions for $K(T-1)$ and $P(T-1)$ are

$$\begin{aligned}K'(T-1) = {} & -[G'(T-1)Q(T)G(T-1) \\ & + R(T)]^{-1}G'(T-1)Q(T)F(T-1)\end{aligned} \tag{14.1-6}$$

$$\begin{aligned}P(T-1) = {} & F'(T-1)\{Q(T) - Q(T)G(T-1)[G'(T-1)Q(T)G(T-1) \\ & + R(T)]^{-1}G'(T-1)Q(T)\}F(T-1).\end{aligned} \tag{14.1-7}$$

We now turn to the calculation of the matrices $K(t)$, determining the optimal control law, and $P(t)$, determining the optimal performance index, for arbitrary values of t. As part of the inductive hypothesis, we assume that $V^*[x(t+1), t+1] = x'(t+1)P(t+1)x(t+1)$ for a certain matrix $P(t+1)$. By proving that $V^*[x(t), t]$ is of the form $x'(t)P(t)x(t)$ for a certain $P(t)$, we will have established the quadratic nature of the performance index. [Of course, the expression for $V^*(x(T-1), T-1)$ derived in Eq. (14.1-5) serves as the first step in the induction.]

Applying the inductive hypothesis to (14.1-3), we have

$$\begin{aligned}V^*[x(t), t] = \min_{u(t)} \{ & u'(t)[G'(t)Q(t+1)G(t) + R(t+1)]u(t) \\ & + 2x'(t)F'(t)Q(t+1)G(t)u(t) + x'(t)F'(t)Q(t+1)F(t)x(t) \\ & + x'(t)F'(t)P(t+1)F(t)x(t) + 2x'(t)F'(t)P(t+1)G(t)u(t) \\ & + u'(t)G'(t)P(t+1)G(t)u(t)\}.\end{aligned}$$

Again, the minimizing $u(t)$, which is the optimal control at time t, is a linear function of $x(t)$,

$$u^*(t) = K'(t)x(t) \tag{14.1-8}$$

and the optimal performance index $V^*(x(t), t)$, resulting from use of $u^*(t)$,

is quadratic in $x(t)$—i.e.,

$$V^*(x(t), t) = x'(t)P(t)x(t). \tag{14.1-9}$$

The expression for $K'(t)$ is

$$\begin{aligned} K'(t) = &-[G'(t)Q(t+1)G(t) + R(t+1) \\ &+ G'(t)P(t+1)G(t)]^{-1}[G(t)Q(t+1)F(t) + G'(t)P(t+1)F(t)] \\ = &-[G'(t)\hat{Q}(t+1)G(t) + R(t+1)]^{-1}G'(t)\hat{Q}(t+1)F(t) \end{aligned} \tag{14.1-10}$$

where

$$\hat{Q}(t+1) = Q(t+1) + P(t+1). \tag{14.1-11}$$

The expression for $P(t)$ is

$$\begin{aligned} P(t) = F'(t)\{&\hat{Q}(t+1) - \hat{Q}(t+1)G(t)[G'(t)\hat{Q}(t+1)G(t) \\ &+ R(t)]^{-1}G'(t)\hat{Q}(t+1)\}F(t). \end{aligned} \tag{14.1-12}$$

Observe that (14.1-11) and (14.1-12) together allow recursive determination of $P(t)$—i.e., determination in sequence and starting with $P(T-1)$ as given in (14.1-7) of $P(T-2), P(T-3), \ldots$. Equation (14.1-10) expresses the optimal feedback law in terms of known quantities, and the members of the sequence $P(T-1), P(T-2), \ldots$.

Thus, to solve a discrete time optimization problem of the sort posed, we must compute the matrices $P(t)$ by starting at the endpoint and working backward in time; from these, the optimal gains must be found.

We shall now consider the infinite-time problem, obtained by letting the final time T in the performance index (14.1-2) go to infinity. The performance index then becomes

$$\begin{aligned} V(x(t_0), u(\cdot), t_0) = \lim_{T\to\infty} \sum_{t=t_0+1}^{T} &[x'(t)Q(t)x(t) \\ &+ u'(t-1)R(t)u(t-1)]. \end{aligned} \tag{14.1-13}$$

To guarantee that the optimal performance index is finite, we shall require this assumption.

> Assumption 14.1-2. For each t, the pair $[F(t), G(t)]$ is completely controllable—i.e., for any state $x(t)$ there exists a time $t_1 > t$ and a control defined on $[t, t_1]$ with the property that application of the control leads to $x(t_1) = 0$.

This assumption is not necessary to guarantee finiteness of the optimal performance index, but it may be shown to be sufficient (see Problem 14.1-1).

In the same way as was done for continuous time systems, we define $x'(t)P(t, T)x(t)$ to be the optimal performance index associated with (14.1-1) and (14.1-2); it is then possible to show that $P(t, T)$ is monotonically increasing with T, and bounded above for all T. We conclude the existence of

$$\lim_{T\to\infty} P(t, T) = \bar{P}(t). \tag{14.1-14}$$

Moreover, it is easily shown that

$$V^*[x(t), t] = x'(t)\bar{P}(t)x(t) \tag{14.1-15}$$

when (14.1-13) is the performance index. Furthermore, $\bar{P}(\cdot)$ satisfies the recursion relation (14.1-12)—i.e., knowing $\bar{P}(t_1)$ for any t_1, we can use (14.1-12) to compute successively $\bar{P}(t_1 - 1)$, $\bar{P}(t_1 - 2), \ldots$, and the optimal control law is given by (14.1-10).

For the case of constant F, G, Q, and R, $\bar{P}(t)$ becomes independent of t. We may then compute $\bar{P}$ via the formula

$$\lim_{t \to -\infty} P(t, T) = \bar{P} \tag{14.1-16}$$

using relation (14.1-12) recursively. [Equation (14.1-16) actually provides a more convenient approach to the evaluation of $\bar{P}$ than does Eq. (14.1-14).] The matrix $\bar{P}$ is also a solution of the "steady state" version of (14.1-12)—i.e.,

$$\bar{P} = F'\{(Q + \bar{P}) - (Q + \bar{P})G[G'(Q + \bar{P})G + R]^{-1}G'(Q + \bar{P})\}F. \tag{14.1-17}$$

However, this equation does not lend itself to the evaluation of $\bar{P}$ in the same way as the corresponding continuous time equation.

The control law in this case also becomes constant, and we see from (14.1-10) that it is given by

$$K' = -[G'(Q + \bar{P})G + R]^{-1}G'(Q + \bar{P})F \tag{14.1-18}$$

with the associated closed-loop system being

$$x(t + 1) = (F + GK')x(t). \tag{14.1-19}$$

As for the continuous time case, an assumption guaranteeing asymptotic stability of the closed-loop system is as follows:

ASSUMPTION 14.1-3. With D any matrix such that $DD' = Q$, the pair $[F, D]$ is completely observable, or, equivalently, the equation $D'F^k x_0 = 0$ for all k implies $x_0 = 0$.

Many of the remarks and extensions applicable to continuous time problems carry over to the discrete time case. For example, one may pose a tracking problem [aimed at finding a control to make the system state $x(\cdot)$ track a desired trajectory $\tilde{x}(\cdot)$]. This is done by using a performance index of the form

$$V(x(t_0), u(\cdot), t_0) = \sum_{t=t_0+1}^{T} \{[x(t) - \tilde{x}(t)]'Q(t)[x(t) - \tilde{x}(t)] + u'(t-1)R(t)u(t-1)\}. \tag{14.1-20}$$

The optimal control turns out, as in the continuous time case, to consist of a linear feedback term and an externally applied term. (See Problem 14.1-5.)

Two primary references to the quadratic minimization problem for discrete time systems are [1] and [2]. More recently, lengthy discussions on this and related matters have appeared in [3] and [4]. For convenience, we shall summarize the principal results.

Finite time regulator. Consider the system

$$x(t+1) = F(t)x(t) + G(t)u(t) \qquad x(t_0) \quad \text{given.} \tag{14.1-1}$$

Let $Q(t)$ and $R(t)$ be nonnegative definite matrices for all t, with $G'(t-1)Q(t)G(t-1) + R(t)$ nonsingular for all t. Define the performance index

$$V(x(t_0), u(\cdot), t_0) = \sum_{t=t_0+1}^{T} [x'(t)Q(t)x(t) + u'(t-1)R(t)u(t-1)]. \tag{14.1-2}$$

Then the minimum value of the performance index is

$$V^*(x(t_0), t_0) = x'(t_0)P(t_0, T)x(t_0)$$

where $P(t, T)$ is defined recursively via

$$\begin{aligned} P(t, T) = F'(t)\{\hat{Q}(t+1) - \hat{Q}(t+1)G(t)[G'(t)\hat{Q}(t+1)G(t) \\ + R(t+1)]^{-1}G'(t)\hat{Q}(t+1)\}F(t) \quad P(T, T) = 0 \end{aligned} \tag{14.1-12}$$

and

$$\hat{Q}(t+1) = Q(t+1) + P(t+1, T). \tag{14.1-11}$$

The associated optimal control law is given by

$$\begin{aligned} u^*(t) = -[G'(t)\hat{Q}(t+1)G(t) \\ + R(t+1)]^{-1}G'(t)\hat{Q}(t+1)F(t). \end{aligned} \tag{14.1-10}$$

Infinite-time regulator. Assuming now that $T \to \infty$, and that the pair $[F(t), G(t)]$ is completely controllable for all t, then

$$\bar{P}(t) = \lim_{T\to\infty} P(t, T) \tag{14.1-14}$$

exists, and the optimum value of the performance index (14.1-2), with T replaced by infinity, is $x'(t_0)\bar{P}(t_0)x(t_0)$. The matrix $\bar{P}$ satisfies the recursion relations (14.1-12) with

$$\hat{Q}(t+1) = Q(t+1) + \bar{P}(t+1) \tag{14.1-21}$$

and the optimal control law is given by (14.1-10).

Time-invariant regulator. When F, G, Q, and R are constant, $\bar{P}$ is constant and may be obtained via

$$\bar{P} = \lim_{t\to-\infty} P(t, T). \tag{14.1-16}$$

The optimal control law is also constant, being given by

$$u^* = -[G'(Q + \bar{P})G + R]^{-1}G'(Q + \bar{P})F. \tag{14.1-18}$$

With D any matrix such that $DD' = Q$, complete observability of the pair $[F, D]$ is sufficient to guarantee asymptotic stability of the closed-loop system.

Problem 14.1-1. For the case of constant F, G, Q, and R with Q positive definite, show that a necessary and sufficient condition for the optimal performance index associated with (14.1-13) to be finite, assuming the optimal index exists, is that no uncontrollable state fails to be asymptotically stable.

Problem 14.1-2. Find the optimal control law and optimal performance indices for the system

$$x(t + 1) = x(t) + u(t) \qquad x(0) \quad \text{given}$$

[where $x(\cdot)$ and $u(\cdot)$ are scalar quantities] with the two performance indices

$$\sum_{t=1}^{3} [2x^2(t) + u^2(t - 1)]$$

and

$$\sum_{t=1}^{\infty} [2x^2(t) + u^2(t - 1)].$$

Problem 14.1-3. When Assumption 14.1-1 fails—i.e., $G'(t - 1)Q(t)G(t - 1) + R(t)$ is singular for some t—in the case of single-input systems this quantity is zero. Now the inverse of this quantity occurs in the formulas for $K'(T - 1)$ and $P(T - 1)$ in (14.1-6) and (14.1-7). Show, by examining the derivation of the optimal control $u^*(T - 1)$, that if $G'(T - 1)Q(T)G(T - 1) + R(T)$ is zero, $u^*(T - 1) = 0$ will be a value, but not the unique value, of the optimal control, and that the inverse may be replaced by zero. Show that for $t < T - 1$, if $G'(t - 1)Q(t)G(t - 1) + G'(t - 1)P(t)G(t - 1) + R(t)$ is zero, the inverse of this quantity in the expressions for $K(t)$ and $P(t)$ may be replaced by zero.

Problem 14.1-4. Given an n-dimensional system, show that the optimal control minimizing the performance index

$$V(x(t_0), u(\cdot), t_0) = x'(t_0 + n)Qx(t_0 + n),$$

where Q is positive definite, will result in a deadbeat ($x(t_0 + n) = 0$) response.

Problem 14.1-5. Given the system

$$x(t + 1) = F(t)x(t) + G(t)u(t) \qquad x(t_0) \quad \text{given}$$

and the performance index

$$V(x(t_0), u(\cdot), t_0) = \sum_{t=t_0+1}^{T} \{[x(t) - \tilde{x}(t)]'Q(t)[x(t) - \tilde{x}(t)] + u'(t - 1)R(t)u(t - 1)\}$$

where $\tilde{x}(\cdot)$ is a sequence of states (the desired system trajectory), find the optimal control.

14.2 THE INFINITE-TIME REGULATOR PROBLEM FOR TIME-VARYING SYSTEMS

The purpose of this section is to present some advanced results concerning the infinite-time regulator problem associated with continuous time systems of the form

$$\dot{x} = F(t)x + G(t)u. \tag{14.2-1}$$

In earlier chapters, we generally restricted discussion of infinite-time problems to time-invariant systems, and merely alluded to the existence of corresponding results for time-varying systems. The results presented here serve to fill in some more details for time-varying systems, but do not appear to have the same practical significance as the corresponding time-invariant results. The primary reference is [5].

The performance index associated with (14.2-1) we take to be

$$V(x(t_0), u(\cdot), t_0) = \lim_{T\to\infty} \int_{t_0}^{T} [u'(t)R(t)u(t) + x'(t)Q(t)x(t)]\, dt \tag{14.2-2}$$

where, as usual, $R(t)$ is positive definite symmetric and $Q(t)$ is nonnegative definite symmetric. We introduce the following assumption:

ASSUMPTION 14.2-1. The pair $[F(t), G(t)]$ is completely controllable for each t.

As we know, it is then possible to write down the optimal control law and optimal performance index in the following fashion. We define $P(t, T)$ to be the solution of

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \qquad P(T, T) = 0 \tag{14.2-3}$$

and we set

$$\bar{P}(t) = \lim_{T\to\infty} P(t, T). \tag{14.2-4}$$

(The existence of the limit may be established.) The optimal control law is

$$u^*(t) = K'(t)x(t) = -R^{-1}(t)G'(t)\bar{P}(t)x(t), \tag{14.2-5}$$

which yields the closed-loop system

$$\dot{x} = (F + GK')x = (F - GR^{-1}G'\bar{P})x. \tag{14.2-6}$$

The optimal performance index is

$$V^*[x(t_0), t_0] = x'(t_0)\bar{P}(t_0)x(t_0). \tag{14.2-7}$$

Questions such as the following then arise:

1. When do $K(t)$ and $\bar{P}(t)$ have bounded entries for all t?
2. When is the closed-loop system of Eq. (14.2-6) asymptotically stable?

We shall now list sufficient conditions (without proof, but with side remarks) for the boundedness of $K(t)$, $\bar{P}(t)$, and for a special form of asymptotic stability of the closed-loop system. This special form of asymptotic stability is *exponential asymptotic stability:* $\dot{x} = F(t)x$ is termed exponentially asymptotically stable if for all t, t_0, and $x(t_0)$, there exist positive constants α_1 and α_2 such that

$$x'(t)x(t) \leq \alpha_1 \exp\left[-2\alpha_2(t - t_0)\right]x'(t_0)x(t_0). \tag{14.2-8}$$

As is evident from this definition, the size of $x(t)$ is bounded above by an exponentially decaying quantity, the initial value of which depends on $x(t_0)$.

There are a number of conditions that, when taken together, will guarantee the results we want. These occur here as Assumptions 14.2-2, 14.2-3, and 14.2-4, the set of conditions having a natural subdivision. We start with the simplest conditions.

Assumption 14.2-2. The following matrices have entries that are bounded for all t: $F(t)$, $G(t)$, $Q(t)$, $R(t)$, and $R^{-1}(t)$.

In common with Assumptions 14.2-3 and 14.2-4, Assumption 14.2-2 may not be necessary to guarantee the desired results in some situations. However, it is perhaps interesting to see how the conditions inherent in it may be relevant. For example, suppose $F(t)$ and $G(t)$ are constant, and $Q(t) = \hat{Q}e^{-2\alpha t}$, $R(t) = \hat{R}e^{-2\alpha t}$, where $\hat{Q}$ is constant and nonnegative definite, $\hat{R}$ is constant and positive definite, and α is a positive constant. We have an optimization problem which is readily convertible to a time-invariant problem. Indeed, the minimization of (14.2-2), given (14.2-1), is equivalent under the identifications

$$\hat{x} = e^{-\alpha t}x$$
$$\hat{u} = e^{-\alpha t}u$$

to the minimization of

$$\int_{t_0}^{\infty} (\hat{u}'\hat{R}\hat{u} + \hat{x}'\hat{Q}\hat{x})\, dt$$

subject to

$$\dot{\hat{x}} = (F - \alpha I)\hat{x} + G\hat{u}.$$

With $[F, G]$ completely controllable and $\hat{Q}$ positive definite, an optimal control law $\hat{u} = K'\hat{x}$ exists with K constant, and with $F - \alpha I + GK'$ possessing eigenvalues all in $\mathrm{Re}[s] < 0$. For the original problem, the optimal control law is $u = K'x$, but now the closed-loop system, $\dot{x} = (F + GK')x$, may not be asymptotically stable!

Thus, in this example are satisfied all the right sort of conditions that we

would expect would guarantee asymptotic stability of the closed-loop system, except for the boundedness of the entries of $R^{-1}(t)$. Yet asymptotic stability may well not be present.

The second condition introduces a new term, defined momentarily.

ASSUMPTION 14.2-3. The pair $[F(t), G(t)]$ is *uniformly completely controllable* [5].

Uniform complete controllability is an extension of the concept of complete controllability. Before we give a quantitative statement of the definition, the following qualitative remarks may help. We recall that a system is completely controllable at time t if, given an arbitrary state $x(t)$ at time t, there exists a control and a time $t_1 > t$ such that application of the control to the system, assumed in state $x(t)$ at time t, will lead to $x(t_1) = 0$. Uniform complete controllability requires, first, that there be complete controllability at every time t, and second, that no matter what t is, the time $t_1 - t$ taken to get to the zero state should be bounded above. Third, the control energy required to transfer the state from $x(t)$ to $x(t_1) = 0$, as measured by

$$\int_t^{t_1} u'(\tau)u(\tau)\, d\tau,$$

should be bounded above and below by positive constants independent of t, and depending only on $x'(t)x(t)$, or the size of $x(t)$. That is, for all t and $x(t)$, there should exist a control $u(\cdot)$ achieving the required state transfer, and such that

$$\alpha_3 x'(t)x(t) \leq \int_t^{t_1} u'(\tau)u(\tau)\, d\tau \leq \alpha_4 x'(t)x(t)$$

for some positive constants α_3 and α_4.

However, this is not all. A system is *completely reachable* at time t if, given an arbitrary state $x(t)$, there exists $t_2 < t$ and a control $u(\cdot)$ such that with $x(t_2) = 0$, the control $u(\cdot)$ will take the system state to $x(t)$ at time t. Three more qualifications of uniformly completely controllable systems are that they are completely reachable at every time t; that no matter what t is, the time $t - t_2$ is bounded above; and that the control energy effecting the transfer from $x(t_2) = 0$ to $x(t)$ should be bounded above and below by positive constants independent of t, and depending only on $x'(t)x(t)$.

A moment's reflection will show that with F and G constant, complete controllability of $[F, G]$ implies uniform complete controllability. {For examples of completely controllable but not uniformly completely controllable pairs $[F(t), G(t)]$, see, e.g., [6].}

The precise quantitative definition of uniform complete controllability is as follows. Let $\Phi(\cdot, \cdot)$ be the transition matrix associated with $\dot{x} = F(t)x$, and define

$$W(t_0, t_1) = \int_{t_0}^{t_1} \Phi(t_0, t)G(t)G'(t)\Phi'(t_0, t)\, dt. \tag{14.2-9}$$

Then the pair $[F(t), G(t)]$ is uniformly completely controllable if two of the following three conditions hold (if any two hold, the third automatically holds):

$$0 < \alpha_3(\sigma)I \leq W(t, t+\sigma) \leq \alpha_4(\sigma)I < \infty, \tag{14.2-10}$$

$$\begin{aligned} 0 < \alpha_5(\sigma)I &\leq \Phi(t+\sigma, t)W(t, t+\sigma)\Phi'(t+\sigma, t) \\ &\leq \alpha_6(\sigma)I < \infty, \end{aligned} \tag{14.2-11}$$

and

$$\max_{i,j} |(\Phi(t, \tau))_{ij}| \leq \alpha_7(|t-\tau|) \qquad \text{for all } t, \tau. \tag{14.2-12}$$

In these equations, $\alpha_3(\cdot), \ldots, \alpha_7(\cdot)$ are continuous functions of their arguments.

One property of uniformly completely controllable pairs that proves useful is the following (see [6]).

Lemma. If $F(t)$ and $G(t)$ have bounded entries, and if the pair $[F(t), G(t)]$ is uniformly completely controllable, then for any $K'(t)$ such that $G(t)K'(t)$ has the same dimension as $F(t)$ and such that the entries of $K(t)$ are bounded, the pair $[F(t) + G(t)K'(t), G(t)]$ is uniformly completely controllable.

It follows from this result, for example, that the commonly occurring pair

$$F(t) = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \\ -a_1(t) & -a_2(t) & -a_3(t) & \cdot & \cdot & -a_n(t) \end{bmatrix} \qquad G(t) = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \tag{14.2-13}$$

is uniformly completely controllable so long as the $a_i(t)$ are all bounded.

The third assumption required for the desired stability results is the following.

Assumption 14.2-4. With $D(t)$ any matrix such that $D(t)D'(t) = Q(t)$, the pair $[F(t), D(t)]$ is *uniformly completely observable*.

As might be imagined, the term uniform complete observability can most readily be defined as the dual of uniform complete controllability: $[F(t), D(t)]$ is uniformly completely observable if and only if $[F'(t), D(t)]$ is uniformly completely controllable.

Assumptions 14.2-2 through 14.2-4 guarantee the following results for the infinite-time regulator problem [5]:

1. The optimal control law gain $K(t)$ has bounded entries for all t.
2. The matrix $\bar{P}(t)$ satisfies

$$\alpha_8 I \leq \bar{P}(t) \leq \alpha_9 I \tag{14.2-14}$$

for all t and for some positive constants α_8 and α_9.
3. The optimal closed-loop system [see Eq. (14.2-6)] is exponentially asymptotically stable, and $x'(t)\bar{P}(t)x(t)$ is a Lyapunov function establishing this stability property.

This completes our discussion of the optimal regulator. We now turn to a related topic, the infinite-time Kalman–Bucy filter. We recall that we consider a plant

$$\dot{x}(t) = F(t)x(t) + G(t)u(t) + v(t) \tag{14.2-15}$$

$$y(t) = H'(t)x(t) + w(t) \tag{14.2-16}$$

where $E[v(t)v'(\tau)] = Q(t)\delta(t-\tau)$, $E[w(t)w'(\tau)] = R(t)\delta(t-\tau)$, with $v(\cdot)$ and $w(\cdot)$ independent zero mean gaussian processes. Furthermore, $E[x(t_0)x'(t_0)] = P_0$, $E[x(t_0)] = m$, and $x(t_0)$ is independent of $v(t)$ and $w(t)$ for all t. Finally, $x(t_0)$ is a gaussian random variable.

The state vector $x(t)$ may be estimated, via an estimate $x_e(t)$, available at time t. The equation describing the calculation of $x_e(t)$ is

$$\begin{aligned}\frac{d}{dt}x_e(t) = [F(t) &+ K_e(t)H'(t)]x_e(t) - K_e(t)y(t) \\ &+ G(t)u(t) \qquad x_e(t_0) = m\end{aligned} \tag{14.2-17}$$

and the matrix $K_e(t)$ is found as follows. Let $P(t)$ be the solution of

$$\dot{P} = PF' + FP - PHR^{-1}H'P + Q \qquad P(t_0) = P_0. \tag{14.2-18}$$

Then

$$K_e(t) = -P(t)H(t)R^{-1}(t). \tag{14.2-19}$$

Equation (14.2-17) describes the optimal filter, concerning which we ask the following question: Under what conditions is the optimal filter asymptotically stable? In effect, we can present only sufficient conditions for exponential asymptotic stability. Although it is possible to derive them by studying an equivalent regulator problem, as we did in Chapter 8, we shall content ourselves here with merely stating the following result. Sufficient conditions for the optimal filter to be exponentially asymptotically stable are [7] as follows:

1. The entries of $F(t)$, $H(t)$, $Q(t)$, $R(t)$, and $R^{-1}(t)$ are bounded for $t \geq t_0$ (or for all t if $t_0 = -\infty$).
2. The pair $[F(t), H(t)]$ is uniformly completely observable.
3. With $D(t)$ any matrix such that $D(t)D'(t) = Q(t)$, the pair $[F(t), D(t)]$ is uniformly completely controllable.

Problem 14.2-1. Prove that if $[F(t), G(t)]$ is uniformly completely controllable, so is $[F(t) + \alpha I, G(t)]$ where α is a positive constant. Suppose that $F(t)$, $G(t)$, $Q(t)$, and $R(t)$ satisfy Assumptions (14.2-2) through (14.2-4) with, as usual, $Q(t)$ nonnegative definite and $R(t)$ positive definite. Show that the performance index

$$\int_{t_0}^{\infty} e^{2\alpha t}[u'(t)R(t)u(t) + x'(t)Q(t)x(t)]\, dt$$

may be minimized by a control law of the form $u(t) = K'(t)x(t)$ with the entries of $K(t)$ bounded, and that the closed-loop system has degree of stability α.

REFERENCES

[1] Kalman, R. E., and T. S. Englar, "A User's Manual for ASP C," *RIAS Rep. NASA*, 1965, Baltimore, Md.

[2] Kalman, R. E., and R. W. Koepcke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indices," *Trans. ASME*, Vol. 80, November 1958, pp. 1820–1826.

[3] Aoki, M., *Optimization of Stochastic Systems*, Academic Press Inc., New York, 1967.

[4] Moore, J. B., and P. M. Colebatch, "Extensions of Quadratic Minimization Theory: The Discrete Time Case," *Proc. IEE*, Vol. 117, No. 1, January 1970, pp. 219–222.

[5] Kalman, R. E., "Contributions to the Theory of Optimal Control," *Bol. Soc. Matem. Mex.*, 1960, pp. 102–119.

[6] Silverman, L. M., and B. D. O. Anderson, "Controllability, Observability and Stability of Linear Systems," *SIAM J. Control*, Vol. 6, No. 1, 1968, pp. 121–130.

[7] Kalman, R. E., and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Trans. ASME Ser. D: J. Basic Eng.*, Vol. 83, March 1961, pp. 95–108.

PART V

COMPUTATIONAL ASPECTS

CHAPTER 15

SOLUTION OF THE RICCATI EQUATION

15.1 PRELIMINARY REMARKS

The selection of an optimal control law and the design of optimal filters require the solving of a Riccati equation. The natural question arises as to how this solution procedure might be carried out. In this chapter, we present a number of different approaches to the solution computation, some of which are only applicable in restricted situations.

It proves convenient to make several subdivisions of the problem of finding Riccati equation solutions, the basis for subdividing being the nature of the coefficient matrices of the Riccati equation and the time interval considered. Throughout this chapter, we restrict attention to the Riccati equation occurring in the control problem, and thus to a situation when a boundary condition will be given at t_1 with the solution desired for $t \leq t_1$. The subdivisions are, then:

1. Time-varying coefficients; $t_1 < \infty$.
2. Time-varying coefficients; $t_1 = \infty$.
3. Constant coefficients; $t_1 < \infty$.
4. Constant coefficients; $t_1 = \infty$.

In all cases, it is possible to solve the Riccati differential equation directly by appropriately programming a digital (or, for that matter, analog) computer. This would appear to be a common procedure, despite the availability of

other procedures to be discussed in later sections. Solution via the digital computer this way naturally requires discretization of time.

When $t_1 = \infty$ and direct solution of the Riccati differential equation is employed, cases 2 and 4 will be treated differently. For case 2, where the coefficients are time varying, we would use the fact that the solution $\bar{P}(t)$ of the Riccati equation for $t_1 = \infty$ is the limit of the solutions $P(t, t_1)$ as t_1 approaches ∞. Therefore, $\bar{P}(t)$ will be obtained approximately by choosing a large t_1, and the approximation will become better as t_1 is increased. In case 4, where $\bar{P}(t)$ is a constant, the fact that the equation coefficients are constant guarantees that

$$\bar{P}(t) = \lim_{t_1 \to \infty} P(t, t_1) = \lim_{t \to -\infty} P(t, t_1). \tag{15.1-1}$$

Therefore, an arbitrary t_1 can be chosen, and the Riccati differential equation solution can be computed backward in time until a constant value is reached.

Whenever computation of the solution is required over a large interval, the question must be considered as to how computational errors introduced at one point in the calculation will propagate. Reference [1] shows that if the closed-loop system resulting from application of an optimal control law is asymptotically stable, then the Riccati differential equation whose solution is used to define the optimal control law is computationally stable. That is, as an error propagates, it is attenuated, and there is no possibility of buildup of the error. We have earlier given conditions for asymptotic stability of the closed-loop system in the time-invariant case; more complex conditions apply in the time-varying case, as discussed in [1] and summarized in Chapter 14.

Tied up with the question of the propagation of computational errors are the infinite-time problem questions of how large t_1 should be taken so that $P(t, t_1)$ will be a good approximation to $\bar{P}(t)$, and how small t should be taken in the constant coefficient matrix case so that $P(t, t_1)$ will be a good approximation to $\bar{P}$. A rough rule of thumb may be applied if the closed-loop system is asymptotically stable: In the constant coefficient case, $t_1 - t$ should be several times the dominant time constant of the closed-loop system, and an appropriate modification of this statement holds in the time-varying coefficient case. Since this time constant will not be known a priori, this rule of thumb is perhaps not particularly helpful, but it is perhaps helpful to know a priori that the sort of interval lengths $t_1 - t$ involved are not long compared with other times that may naturally be of interest.

Other ways of solving the Riccati equation will be discussed in the next two sections. In all four cases 1 through 4, it is possible to replace the problem of solving the *nonlinear* Riccati differential equation by the problem of solving a *linear* differential equation, and then computing a matrix inverse. This is discussed in the next section. The third section discusses a particular approach applying only to case 4. In this case, it is known that the Riccati equation solution is constant and satisfies an algebraic rather than a differ-

ential equation; direct solution procedures for the algebraic equation can then be used.

The basic reference discussing the use of linear differential equations for solving Riccati equations is [2]. The theory behind this computational technique is given in [2], and earlier in, for example, [3] and [4], with simplifications in the constant coefficient matrix case appearing in [5] and [6]. Techniques aimed at solving the steady state algebraic equation have been given in [7] through [11]. A recent survey comparing various techniques (and favoring a technique based on a result in [9]) may be found in [12].

In the fourth section of this chapter, we consider approximate solution techniques for high-order Riccati equations, which rely on the solving of a low-order Riccati equation and linear equations.

15.2 RICCATI EQUATION SOLUTION WITH THE AID OF A LINEAR DIFFERENTIAL EQUATION

For the moment, we restrict attention to the finite-interval problem. Thus, we suppose we are given the Riccati equation

$$-\frac{dP(t)}{dt} = P(t)F(t) + F'(t)P(t) - P(t)G(t)R^{-1}(t)G'(t)P(t) + Q(t)$$

$$P(t_1) = A. \tag{15.2-1}$$

We shall establish the following result (see, e.g., [2] through [4]).

Constructive procedure for obtaining P(t). Consider the equations

$$\frac{d}{dt}\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} F & -GR^{-1}G' \\ -Q & -F' \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix}; \quad \begin{bmatrix} X(t_1) \\ Y(t_1) \end{bmatrix} = \begin{bmatrix} I \\ A \end{bmatrix} \tag{15.2-2}$$

where the explicit dependence of the matrices on t is suppressed. Then, provided that the solution of (15.2-1) exists on the interval $[t, t_1]$, the solution of (15.2-2) has the property that $X^{-1}(t)$ exists and that

$$P(t) = Y(t)X^{-1}(t). \tag{15.2-3}$$

Conversely, the solution of (15.2-1) exists on $[t, t_1]$ if the solution of (15.2-2) has the property that $X(\sigma)$ is nonsingular for all σ in $t \leq \sigma \leq t_1$. The solution of (15.2-1) is given by (15.2-3).

In the usual application of the Riccati equation to designing optimal regulators, the existence of $P(t)$ for all $t \leq t_1$ is assured. [Recall that sufficient conditions for this are as follows: $Q(t)$ is nonnegative definite symmetric for all t, $R(t)$ is positive definite symmetric for all t, and A is nonnegative definite symmetric.] Accordingly, we can normally forget about the existence remarks

in the preceding constructive procedure and concentrate on the actual construction. As is evident from the statement of the procedure, basically two operations are required.

1. Solution of a linear differential equation, of matrix dimension twice that of the dimension of the matrix $F(t)$.
2. Matrix inversion and multiplication.

We shall now verify the procedure. We do this by showing that the matrix $Y(t)X^{-1}(t)$ satisfies the Riccati differential equation (15.2-1). It then follows, by the uniqueness property of solutions of differential equations with continuous coefficients, that $P(t)$ is given by (15.2-3). Formally,

$$\begin{aligned}\frac{d}{dt}[YX^{-1}] &= \dot{Y}X^{-1} - YX^{-1}\dot{X}X^{-1} \\ &= -QXX^{-1} - F'YX^{-1} - YX^{-1}FXX^{-1} \\ &\quad + YX^{-1}GR^{-1}G'YX^{-1}\end{aligned}$$

on using (15.2-2). Therefore,

$$-\frac{d}{dt}[YX^{-1}] = [YX^{-1}]F + F'[YX^{-1}] - [YX^{-1}]GR^{-1}G'[YX^{-1}] + Q.$$

This is the differential equation (15.2-1). Note, also, that the boundary condition in (15.2-2) yields immediately that $Y(t_1)X^{-1}(t_1) = A$, which is the required boundary condition. Consequently, (15.2-3) is established.

The preceding manipulations also show that if $X^{-1}(\sigma)$ exists for all σ in $[t, t_1]$, a solution of (15.2-1) exists over the same interval. Thus, the second of the two existence claims in the constructive procedure statement is verified.

Let us now check that existence of a solution $P(t)$ to (15.2-1) guarantees that $X^{-1}(t)$ exists. This will complete the proof of all claims made in the constructive procedure statement.

Let $\Phi(\cdot, \cdot)$ be the transition matrix associated with

$$\frac{dx}{dt} = [F(t) - G(t)R^{-1}(t)G'(t)P(t)]x. \tag{15.2-4}$$

Because $P(t)$ exists for $t \leq t_1$, $\Phi(\cdot, \cdot)$ is defined for all values of its arguments less than or equal to t_1. We claim that

$$X(t) = \Phi(t, t_1) \qquad Y(t) = P(t)\Phi(t, t_1) \tag{15.2-5}$$

satisfy (15.2-2) including its boundary condition. This is sufficient to prove that $X^{-1}(t)$ exists, since $\Phi^{-1}(t, t_1) = \Phi(t_1, t)$ is known to exist.

To verify the claim (15.2-5), we have

$$\begin{aligned}\frac{d}{dt}[\Phi(t, t_1)] &= [F(t) - G(t)R^{-1}(t)G'(t)P(t)]\Phi(t, t_1) \\ &= F(t)[\Phi(t, t_1)] - G(t)R^{-1}(t)G'(t)[P(t)\Phi(t, t_1)].\end{aligned}$$

Also,

$$\begin{aligned}\frac{d}{dt}[P(t)\Phi(t, t_1)] &= \frac{dP(t)}{dt}\Phi(t, t_1) + P(t)\frac{d}{dt}\Phi(t, t_1)\\ &= -P(t)F(t)\Phi(t, t_1) - F'(t)P(t)\Phi(t, t_1)\\ &\quad + P(t)G(t)R^{-1}(t)G'(t)P(t)\Phi(t, t_1) - Q(t)\Phi(t, t_1)\\ &\quad + P(t)F(t)\Phi(t,t_1) - P(t)G(t)R^{-1}(t)G'(t)P(t)\Phi(t, t_1)\\ &= -Q(t)[\Phi(t, t_1)] - F'(t)[P(t)\Phi(t, t_1)].\end{aligned}$$

Therefore, we have verified that, with the identification (15.2-5), the differential equation part of (15.2-2) holds. Also, with the identification (15.2-5) and the boundary condition $P(t_1) = A$, we recover the boundary condition part of (15.2-2). The identifications (15.2-5) are therefore validated.

An alternative expression for $P(t)$ is available in terms of the transition matrix associated with the equations (15.2-2). With F an $n \times n$ matrix, let the $2n \times 2n$ matrix $\Theta(\cdot, \cdot)$ be the transition matrix of (15.2-2)—i.e.,

$$\frac{d}{dt}\Theta(t, \tau) = \begin{bmatrix} F & -GR^{-1}G' \\ -Q & -F' \end{bmatrix}\Theta(t, \tau) \tag{15.2-6}$$

$$\Theta(\tau, \tau) = I_{2n} \qquad \text{for all } \tau. \tag{15.2-7}$$

Partition $\Theta(t, \tau)$ into four $n \times n$ submatrices, as

$$\Theta = \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix}. \tag{15.2-8}$$

Then (15.2-2) implies

$$X(t) = \Theta_{11}(t, t_1) + \Theta_{12}(t, t_1)A$$
$$Y(t) = \Theta_{21}(t, t_1) + \Theta_{22}(t, t_1)A,$$

and so

$$\begin{aligned}P(t) = [\Theta_{21}(t, t_1) + \Theta_{22}(t, t_1)A][\Theta_{11}(t, t_1)\\ + \Theta_{12}(t, t_1)A]^{-1}.\end{aligned} \tag{15.2-9}$$

For computing $P(t)$ when F, G, Q, and R are constant, the advantage of using this procedure may be substantial, since an analytical formula is available for the transition matrix $\Theta(\cdot, \cdot)$. Certainly this formula involves matrix exponentiation, which itself may be hard. But in many instances this may be straightforward, and in many other instances it may be tackled by one of the many numerical techniques now being developed for matrix exponentiation.

To illustrate the foregoing procedure, we shall solve

$$-\dot{P} = P\begin{bmatrix} 0 & 0 \\ 1 & -\sqrt{2} \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & -\sqrt{2} \end{bmatrix}P - P\begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \quad 0]P + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

with $P(0) = 0$. We first construct the equation for Θ:

$$\dot{\Theta} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & -\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & \sqrt{2} \end{bmatrix} \Theta.$$

To carry out the matrix exponentiation required to represent Θ analytically, we shall conduct a similarity transformation on the preceding coefficient matrix to obtain the associated Jordan form. The similarity transformation is constructed by first evaluating the eigenvalues of the coefficient matrix—which turn out to be $+1$ twice and -1 twice—and then finding the associated eigenvectors. There are actually only two proper eigenvectors, and the Jordan form turns out to have two nonzero elements on the superdiagonal. Once this is established, it is easy to compute the complete similarity transformation. We have

$$\begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & -\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & 1 \\ \sqrt{2}-1 & \sqrt{2}-2 & \sqrt{2}+1 & -\sqrt{2}-2 \\ -1 & 0 & 1 & 0 \\ 1 & 1 & 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 4 & 1 \\ \sqrt{2}-1 & \sqrt{2}-2 & \sqrt{2}+1 & \sqrt{2}-2 \\ -1 & 0 & 1 & 0 \\ 1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Writing this equation symbolically as

$$AT = TJ,$$

then

$$e^{Jt} = \begin{bmatrix} e^t & te^t & 0 & 0 \\ 0 & e^t & 0 & 0 \\ 0 & 0 & e^{-t} & te^{-t} \\ 0 & 0 & 0 & e^{-t} \end{bmatrix}$$

and

$$e^{At} = Te^{Jt}T^{-1}.$$

Now, T^{-1} may be computed to be

$$T^{-1} = \tfrac{1}{16}\begin{bmatrix} 4 & 0 & -8 & 4 \\ -4 & -4 & 4 & 4\sqrt{2}+4 \\ 4 & 0 & 8 & 4 \\ 4 & -4 & 4 & 4\sqrt{2}-4 \end{bmatrix}.$$

Therefore,

$$\Theta(t, \tau) = e^{A(t-\tau)}$$

and $\Theta(t) = \frac{1}{4}$ times the matrix,

$$\begin{bmatrix}
\begin{array}{l} 2e^t - te^t \\ +2e^{-t} + te^{-t} \end{array} & \begin{array}{l} e^t - te^t \\ -e^{-t} - te^{-t} \end{array} & \begin{array}{l} -3e^t + te^t \\ +3e^{-t} + te^{-t} \end{array} & \begin{array}{l} \sqrt{2}e^t + (1 + \sqrt{2})te^t \\ +\sqrt{2}e^{-t} + (\sqrt{2} - 1)te^{-t} \end{array} \\[2ex]
\begin{array}{l} e^t - (\sqrt{2} - 1)te^t \\ -e^{-t} + (\sqrt{2} + 1)te^{-t} \end{array} & \begin{array}{l} (2 - \sqrt{2})e^t - (\sqrt{2} - 1)te^t \\ +(2 + \sqrt{2})e^{-t} - (\sqrt{2} + 1)te^{-t} \end{array} & \begin{array}{l} -\sqrt{2}e^t + (\sqrt{2} - 1)te^t \\ +\sqrt{2}e^{-t} + (\sqrt{2} + 1)te^{-t} \end{array} & \begin{array}{l} -e^t + te^t \\ +e^{-t} + te^{-t} \end{array} \\[2ex]
\begin{array}{l} -e^t + te^t \\ +e^{-t} + te^{-t} \end{array} & te^t - te^{-t} & \begin{array}{l} 2e^t - te^t \\ +2e^{-t} + te^{-t} \end{array} & \begin{array}{l} -e^t - (1 + \sqrt{2})te^t \\ +e^{-t} + (\sqrt{2} - 1)te^{-t} \end{array} \\[2ex]
-te^t + te^{-t} & \begin{array}{l} -e^t - te^t \\ +e^{-t} - te^{-t} \end{array} & \begin{array}{l} -e^t + te^t \\ +e^{-t} + te^{-t} \end{array} & \begin{array}{l} (2 + \sqrt{2})e^t + (\sqrt{2} + 1)te^t \\ +(2 - \sqrt{2})e^{-t} + (\sqrt{2} - 1)te^{-t} \end{array}
\end{bmatrix}.$$

By Eq. (15.2-9), it then follows that

$$P(t) = \begin{bmatrix} -e^t + te^t & te^t - te^{-t} \\ +e^{-t} + te^{-t} & \\ & \\ -te^t + te^{-t} & -e^t - te^t \\ & +e^{-t} - te^{-t} \end{bmatrix}$$

$$\times \begin{bmatrix} 2e^t - te^t & e^t - te^t \\ +2e^{-t} + te^{-t} & -e^{-t} + te^{-t} \\ & \\ e^t - (\sqrt{2} - 1)te^t & (2 - \sqrt{2})e^t - (\sqrt{2} - 1)te^t \\ -e^{-t} + (\sqrt{2} + 1)te^{-t} & +(2 + \sqrt{2})e^{-t} - (\sqrt{2} + 1)te^{-t} \end{bmatrix}^{-1}.$$

We can also evaluate $\lim_{t \to -\infty} P(t)$, which should be a steady state solution of the Riccati equation satisfied by P. To do this, we can delete terms involving e^t or te^t in the preceding expression for $P(t)$ as these become negligibly small, so long as the inverse in the formula for $P(t)$ exists after this deletion. Thus,

$$\lim_{t \to -\infty} P(t) = \lim_{t \to -\infty} \begin{bmatrix} e^{-t} + te^{-t} & -te^{-t} \\ te^{-t} & e^{-t} - te^{-t} \end{bmatrix}$$

$$\times \begin{bmatrix} 2e^{-t} + te^{-t} & -e^{-t} - te^{-t} \\ -e^{-t} + (\sqrt{2} + 1)te^{-t} & (2 + \sqrt{2})e^{-t} - (\sqrt{2} + 1)te^{-t} \end{bmatrix}^{-1}. \tag{15.2-10}$$

Indeed, the necessary inverse exists, and proceeding further with the calculation, we have

$$\lim_{t \to -\infty} P(t) = \lim_{t \to -\infty} \begin{bmatrix} e^{-t} + te^{-t} & -te^{-t} \\ te^{-t} & e^{-t} - te^{-t} \end{bmatrix} \frac{1}{(3 + 2\sqrt{2})e^{-2t}}$$

$$\times \begin{bmatrix} (2 + \sqrt{2})e^{-t} - (\sqrt{2} + 1)te^{-t} & e^{-t} + te^{-t} \\ e^{-t} - (\sqrt{2} + 1)te^{-t} & 2e^{-t} + te^{-t} \end{bmatrix}$$

$$= \lim_{t \to -\infty} \frac{1}{(3 + 2\sqrt{2})e^{-2t}} \begin{bmatrix} (2 + \sqrt{2})e^{-2t} & e^{-2t} \\ e^{-2t} & 2e^{-2t} \end{bmatrix}$$

$$= \begin{bmatrix} 2 - \sqrt{2} & 3 - 2\sqrt{2} \\ 3 - 2\sqrt{2} & 6 - 4\sqrt{2} \end{bmatrix}.$$

It is readily verified that this constant matrix both satisfies the Riccati equation and is a positive definite matrix.

To provide an analytically solvable example, we restricted attention to a Riccati equation with constant coefficient matrices. Note, however, that the computation of the Riccati equation solution using a linear differential equation certainly does not require constant coefficient matrices.

References [5] and [6] offer certain comments applicable when the matrix

coefficients are constant. Reference [5] is concerned with setting up formulas for $P(t)$ and the limiting $\bar{P}$ by the same diagonalization procedure as used previously, and with comparing the limiting $\bar{P}$ formulas with those obtainable by other techniques. Reference [6] extends these ideas and removes a computational problem which may arise.

To understand the nature of this problem, consider the preceding example, and, in particular, Eq. (15.2-10). The 1-1 entry of the first matrix on the right side of this equation is $e^{-t} + te^{-t}$, and as $t \to -\infty$, both summands become infinite. But they do so at different rates, and one might be tempted to neglect the e^{-t} summand, or to overlook it in a numerical calculation. Yet to do so would prove fatal, as a major error would be introduced into the final answer. This difficulty is not peculiar to the present example either; in general, terms such as $\alpha_1 e^{-\lambda_1 t} + \alpha_2 e^{-\lambda_2 t}$ will be involved in an expression like (15.2-10), with t approaching minus ∞. Although the term with the greatest value of λ_i will dominate, no other terms may normally be neglected. Accidental neglecting of the other terms—a strong possibility in numerical computation—will lead to a major error in the result.

The difficulty is eliminated in the following way. To solve the equation

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \qquad P(0) = A \tag{15.2-11}$$

with F, G, Q, and R constant and the other usual constraints, we start with the matrix

$$M = \begin{bmatrix} F & -GR^{-1}G' \\ -Q & -F' \end{bmatrix}. \tag{15.2-12}$$

It is possible to show that this matrix has no purely imaginary eigenvalues, and that if λ is an eigenvalue, so is $-\lambda$ (see [5] and Problem 15.2-4). Let

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \tag{15.2-13}$$

be a matrix with the property that

$$T^{-1}MT = \begin{bmatrix} -\Lambda & 0 \\ 0 & \Lambda \end{bmatrix} \tag{15.2-14}$$

where Λ is a matrix that is the direct sum of 1×1 blocks $[\lambda_i]$ with $\lambda_i > 0$, or 2×2 blocks

$$\begin{bmatrix} \lambda_i & \mu_i \\ -\mu_i & \lambda_i \end{bmatrix}$$

with $\lambda_i > 0$, λ_i and μ_i always being real, and Jordan blocks of greater size again. Note that T is also real. Let the $n \times n$ matrices $X(t)$ and $Y(t)$ satisfy

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} = \begin{bmatrix} F & -GR^{-1}G' \\ -Q & -F' \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \tag{15.2.2}$$

or

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} = M \begin{bmatrix} X \\ Y \end{bmatrix}$$

with $X(0) = I$, $Y(0) = A$. Then define new $n \times n$ matrices $\hat{X}(t)$ and $\hat{Y}(t)$ by

$$\begin{bmatrix} X \\ Y \end{bmatrix} = T \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} \tag{15.2-15}$$

It follows that

$$\begin{bmatrix} \dot{\hat{X}} \\ \dot{\hat{Y}} \end{bmatrix} = \begin{bmatrix} -\Lambda & 0 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix}$$

and, therefore,

$$\begin{bmatrix} \hat{X}(0) \\ \hat{Y}(t) \end{bmatrix} = \begin{bmatrix} e^{\Lambda t} & 0 \\ 0 & e^{\Lambda t} \end{bmatrix} \begin{bmatrix} \hat{X}(t) \\ \hat{Y}(0) \end{bmatrix}. \tag{15.2-16}$$

The boundary condition $X(0) = I$, $Y(0) = A$ implies by (15.2-15) that $I = T_{11}\hat{X}(0) + T_{12}(0)\hat{Y}(0)$ and $A = T_{21}\hat{X}(0) + T_{22}\hat{Y}(0)$, or

$$\hat{Y}(0) = R\hat{X}(0) \tag{15.2-17}$$

where

$$R = -[T_{22} - AT_{12}]^{-1}[T_{21} - AT_{11}]. \tag{15.2-18}$$

Now we shall find a new expression for $P(t) = Y(t)X^{-1}(t)$, which we recall is the solution of (15.2-1). Using (15.2-16) and (15.2-17), we obtain immediately the following relation between $\hat{Y}(t)$ and $\hat{X}(t)$:

$$\hat{Y}(t) = e^{\Lambda t}Re^{\Lambda t}\hat{X}(t).$$

Now, using (15.2-15), we obtain

$$P(t) = (T_{21} + T_{22}e^{\Lambda t}Re^{\Lambda t})(T_{11} + T_{12}e^{\Lambda t}Re^{\Lambda t})^{-1} \tag{15.2-19}$$

This is the desired new formula for $P(t)$. A number of points should be noted. First, there are no growing exponentials as t approaches minus infinity appearing in the formula for $P(t)$; in fact, all exponentials decay. Second, when t approaches minus infinity, the limiting $P(t)$ is given simply as

$$\bar{P} = \lim_{t \to -\infty} P(t) = T_{21}T_{11}^{-1}. \tag{15.2-20}$$

(It is possible to show that T_{11} is always nonsingular.) Third, the limiting value of $P(t)$ is independent of R and thus of the initial condition A of the Riccati equation. Fourth, from the point of view of simply computing $P(t)$ rather than studying the theory behind the computation, the steps leading up to (15.2-19) are almost all dispensable. One starts with the matrix M of (15.2-12) and generates the matrix T of (15.2-13), which "almost" diagonalizes M [see (15.2-14)]. With the definition of R in Eq. (15.2-18), the formula (15.2-19) then gives $P(t)$ and, inherently, $\lim_{t \to -\infty} P(t)$. Fifth, the formula (15.2-19)

does not suffer from the computational disadvantage of the formula (15.2-9), upon which we earlier remarked.

In the next section, we discuss more thoroughly the calculation of $\bar{P}$. We shall also indicate an expression for $P(t)$ in terms of $\bar{P}$.

Problem 15.2-1. Consider the Riccati equation

$$-\dot{p} = 2fp - \frac{p^2}{r} + q$$

with initial condition $p(t_1) = 0$. Solve this equation by the technique given in this section and compare your result with that obtained from

$$\int_{p(t)}^{p(t_1)} \frac{dp}{2fp - (p^2/r) + q} = t_1 - t.$$

Problem 15.2-2. Solve the Riccati equation

$$-\dot{P} = P\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} P - \frac{1}{r} P \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1 \quad 0] P$$

$$P(0) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

using the first method outlined in the text.

Problem 15.2-3. Solve the following Riccati equation forward in time, obtaining the limiting solution:

$$\dot{P} = P\begin{bmatrix} 0 & 0 \\ -1 & \sqrt{2} \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 0 & \sqrt{2} \end{bmatrix} P - P \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1 \quad 0] P + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

with boundary condition $P(0) = 0$.

Problem 15.2-4. Consider the matrix M of (15.2-12) with $Q = Q'$ nonnegative definite. Show that if λ is an eigenvalue of M, so is $-\lambda$. Can you show that there are no purely imaginary eigenvalues of M? (*Hint for first part:* From an eigenvector u satisfying $Mu = \lambda u$, construct a row vector v' such that $v'M = -\lambda v'$.)

Problem 15.2-5. Using the method outlined after Eq. (15.2-11), solve the following Riccati equation, which appeared earlier in the worked example:

$$-\dot{P} = P\begin{bmatrix} 0 & 0 \\ 1 & -\sqrt{2} \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & -\sqrt{2} \end{bmatrix} P - P \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1 \quad 0] P + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

with $P(0) = 0$.

15.3 STEADY STATE SOLUTION OF THE RICCATI EQUATION

In this section, we further restrict the Riccati equation problem considered. The coefficient matrices of the equation are constant, and we seek only

the steady state (and constant) solution $\bar{P}$ of the Riccati equation which is positive definite. (It is implicitly assumed throughout the following material that the controllability condition guaranteeing existence of $\bar{P}$ is satisfied, and that the observability condition guaranteeing positive definiteness rather than nonnegative definiteness of the limiting solution is also satisfied.) The actual equation satisfied by $\bar{P}$ is

$$\bar{P}F + F'\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q = 0 \tag{15.3-1}$$

which is an algebraic quadratic matrix equation.

We shall present several different approaches to the solution of (15.3-1). One has already been more or less covered in the previous section, but we shall review it again here. This procedure is best discussed by Potter in [7], although it appears to have been essentially given in [8].

From the coefficient matrices in (15.3-1), we construct the matrix

$$M = \begin{bmatrix} F & -GR^{-1}G' \\ -Q & -F' \end{bmatrix} \tag{15.3-2}$$

which can be shown to have the property that there are no pure imaginary eigenvalues, and the property that if λ is an eigenvalue, so is $-\lambda$. We then construct a matrix

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \tag{15.3-3}$$

that takes M almost to diagonal form, in accordance with the formula

$$T^{-1}MT = \begin{bmatrix} -\Lambda & 0 \\ 0 & \Lambda \end{bmatrix} \tag{15.3-4}$$

where Λ is the direct sum of matrices $[\lambda_i]$ with $\lambda_i > 0$, matrices

$$\begin{bmatrix} \lambda_i & -\mu_i \\ \mu_i & \lambda_i \end{bmatrix}$$

again with λ_i greater than zero, and Jordan blocks of greater size again. Then the desired matrix $\bar{P}$ is given by

$$\bar{P} = T_{21}T_{11}^{-1}. \tag{15.3-5}$$

It is perhaps instructive to show directly that the quadratic equation (15.3-1) is satisfied by (15.3-5). (The direct proof that $\bar{P}$ is positive definite is a little more difficult; since we are primarily interested here in presenting computational procedures, we refer the interested reader to [7]. In any case, the material of the last section provides an indirect proof.) To verify that (15.3-5) is a solution, we have from (15.3-4) that

$$MT = T\begin{bmatrix} -\Lambda & 0 \\ 0 & \Lambda \end{bmatrix}$$

and therefore

$$FT_{11} - GR^{-1}G'T_{21} = -T_{11}\Lambda$$
$$-QT_{11} - F'T_{21} = -T_{21}\Lambda.$$

The first of these yields

$$T_{21}T_{11}^{-1}F - T_{21}T_{11}^{-1}GR^{-1}G'T_{21}T_{11}^{-1} = -T_{21}T_{11}^{-1}T_{11}\Lambda T_{11}^{-1}$$

and the second yields

$$-Q - F'T_{21}T_{11}^{-1} = -T_{21}T_{11}^{-1}T_{11}\Lambda T_{11}^{-1},$$

whence

$$-Q - F'T_{21}T_{11}^{-1} = T_{21}T_{11}^{-1}F - T_{21}T_{11}^{-1}GR^{-1}G'T_{21}T_{11}^{-1}$$

and the desired result follows.

To illustrate the procedure, we consider the steady state equation

$$\bar{P}\begin{bmatrix} 0 & 0 \\ 1 & -\sqrt{2} \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & -\sqrt{2} \end{bmatrix}\bar{P} - \bar{P}\begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \quad 0]\bar{P} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 0.$$

(The associated differential equation was solved as a worked example in the last section.) The matrix M becomes

$$M = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & -\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & \sqrt{2} \end{bmatrix}.$$

The eigenvalues and eigenvectors of M were calculated in the last section. This calculation yields the equation

$$T^{-1}MT = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

with

$$T = \begin{bmatrix} 1 & 1 & 1 & -1 \\ \sqrt{2}+1 & -\sqrt{2}-2 & \sqrt{2}-1 & \sqrt{2}-2 \\ 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 1 \end{bmatrix}.$$

Thus,

$$\begin{aligned}\bar{P} &= \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ \sqrt{2}+1 & -\sqrt{2}-2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 2-\sqrt{2} & 3-2\sqrt{2} \\ 3-2\sqrt{2} & 6-4\sqrt{2} \end{bmatrix}\end{aligned}$$

which agrees with the result of the previous section.

A second procedure for solving (15.3-1), which is closely related to the preceding procedure, is given in [9]. Whereas the previous method requires computation of the eigenvalues of the matrix M and essentially its eigenvectors (which are more or less the columns of the matrix T), the procedure of [9] requires computation of the eigenvalues only. This may represent a saving in computation time, although there are aspects of the calculation which in places make it more unwieldly than the aforementioned procedure.

We state the new procedure first, and then justify it. Starting with the matrix M of (15.3-2), we compute the eigenvalues of M and form that polynomial $p(s)$ whose zeros consist of the *left* half-plane (negative real part) eigenvalues of M. Suppose, for the sake of argument, that

$$p(s) = s^n + a_n s^{n-1} + \cdots + a_1 \tag{15.3-6}$$

which has degree equal to the size of the square matrix F. We next construct the $2n \times 2n$ matrix

$$p(M) = M^n + a_n M^{n-1} + \cdots + a_1 I. \tag{15.3-7}$$

Then $\bar{P}$ is uniquely defined by

$$p(M)\begin{bmatrix} I \\ \bar{P} \end{bmatrix} = 0. \tag{15.3-8}$$

To justify the procedure, let T be the matrix of (15.3-3), which achieves the pseudodiagonalization of (15.3-4). Then

$$\begin{aligned} T^{-1}p(M)T = p(T^{-1}MT) &= \begin{bmatrix} p(-\Lambda) & 0 \\ 0 & p(\Lambda) \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & p(\Lambda) \end{bmatrix}. \end{aligned}$$

The last equality follows because the zeros of $p(s)$ are the negative real part eigenvalues of M, which are also the eigenvalues of $-\Lambda$. The Cayley–Hamilton theorem then guarantees that $p(-\Lambda)$ will equal zero. Now

$$p(M)\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & p(\Lambda) \end{bmatrix}$$

whence

$$p(M)\begin{bmatrix} T_{11} \\ T_{21} \end{bmatrix} = 0$$

or

$$p(M)\begin{bmatrix} I \\ T_{21}T_{11}^{-1} \end{bmatrix} = 0.$$

The method first discussed in this section shows that $T_{21}T_{11}^{-1}$ is precisely $\bar{P}$,

the desired matrix, and hence (15.3-8) is established. The fact that there can only be one matrix X satisfying

$$p(M)\begin{bmatrix} I \\ X \end{bmatrix} = 0$$

(i.e., that $\bar{P}$ is uniquely determined) can also be shown, but we shall omit the proof here because it is not particularly illuminating.

Notice that (15.3-8) is essentially a set of simultaneous linear equations for the elements of $\bar{P}$. Consequently, it is easy to solve. To illustrate the procedure, we take the same example as before—viz.,

$$\bar{P}\begin{bmatrix} 0 & 0 \\ 1 & -\sqrt{2} \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & -\sqrt{2} \end{bmatrix}\bar{P} - \bar{P}\begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \quad 0]\bar{P} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

with

$$M = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & -\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & \sqrt{2} \end{bmatrix}.$$

The eigenvalues of M are -1 twice and $+1$ twice. Accordingly,

$$p(s) = s^2 + 2s + 1$$

and

$$\begin{aligned} p(M) &= M^2 + 2M + I \\ &= \begin{bmatrix} 1 & 0 & -2 & 1 \\ -\sqrt{2} + 2 & -2\sqrt{2} + 3 & -1 & 0 \\ 0 & 1 & 1 & -\sqrt{2} - 2 \\ -1 & -2 & 0 & 2\sqrt{2} + 3 \end{bmatrix}. \end{aligned}$$

Equation (15.3-8) becomes

$$\begin{bmatrix} 1 & 0 & -2 & 1 \\ \sqrt{2} + 2 & -2\sqrt{2} + 3 & -1 & 0 \\ 0 & 1 & 1 & -\sqrt{2} - 2 \\ -1 & -2 & 0 & 2\sqrt{2} + 3 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \bar{p}_{11} & \bar{p}_{12} \\ \bar{p}_{12} & \bar{p}_{22} \end{bmatrix} = 0.$$

This implies eight linear relations among the $\bar{p}_{ij}$, only three of which are independent. Writing down only independent relations, we obtain

$$\begin{aligned} 1 - 2\bar{p}_{11} + \bar{p}_{12} &= 0 \\ -\sqrt{2} + 2 - \bar{p}_{11} &= 0 \\ -2\bar{p}_{12} + \bar{p}_{22} &= 0 \end{aligned}$$

whence

$$\bar{P} = \begin{bmatrix} 2 - \sqrt{2} & 3 - 2\sqrt{2} \\ 3 - 2\sqrt{2} & 6 - 4\sqrt{2} \end{bmatrix}$$

as earlier established.

One characteristic of the two previous methods is that the eigenvalues of a certain matrix must be calculated. This is also a requirement in the method of the last section for computing the differential as distinct from the algebraic equation solution by diagonalizing a matrix to find its exponential.

For large-order systems, eigenvalue computation may pose some problems. Therefore, the next method to be discussed, presented in [10], may provide an acceptable alternative because no eigenvalue computation or polynomial factoring is required. The penalty is that the method recursively computes a sequence of estimates for the positive definite solution of (13.3-1), and, of course, in any given problem many iterations might be required.

As usual, we consider the equation

$$\bar{P}F + F'\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q = 0 \tag{15.3-1}$$

and seek its positive definite solution. The first step in the procedure is to select a matrix K_0 such that $F + GK_0'$ has all its eigenvalues with negative real parts. This is always possible if $[F, G]$ is completely controllable (see [13] through [20]). Then a sequence $P_0, P_1, P_2, \ldots,$ of $n \times n$ matrices is defined in the following fashion. With

$$F_i = F + GK_i' \tag{15.3-9}$$

the matrix P_i satisfies

$$P_iF_i + F'P_i + K_iRK_i' + Q = 0. \tag{15.3-10}$$

Furthermore,

$$K_{i+1} = -P_iGR^{-1} \tag{15.3-11}$$

which permits the carrying through of the next step of the iteration.

It turns out that

$$\lim_{i \to \infty} P_i = \bar{P}. \tag{15.3-12}$$

Evidently, the computation procedure falls into two distinct phases. In the first phase, a matrix K_0 must be determined, such that $F + GK_0'$ has all its eigenvalues with negative real parts. If the eigenvalues of F all have negative real parts, $K_0 = 0$ is an immediate acceptable selection. If not, more computation is required.

In the case when G is a vector rather than a matrix, though, the computation is straightforward. By first making a basis change if necessary, we can assume that F is in companion matrix form. That is,

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \\ -a_1 & -a_2 & -a_3 & \cdot & \cdot & -a_n \end{bmatrix} \tag{15.3-13}$$

and

$$g = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}. \tag{15.3-14}$$

Then, if $k' = [k_1 \quad k_2 \quad \cdots \quad k_n]$, it is immediate that $F + gk'$ is also a companion matrix, and the associated characteristic polynomial is $s^n + (a_n - k_n)s^{n-1} + \cdots + (a_1 - k_1)$. Choice of the k_i to ensure that this polynomial has all its zeros with negative real parts is straightforward. In the case when G is not a vector but a matrix, selection of an appropriate K_0 can be more involved (see [13] through [20]).

In the second phase of the calculations leading to $\bar{P}$, equations such as (15.3-10) are solved. Solution procedures for such equations are well known (see, e.g., [21]); the equation really is a rearranged version of a set of linear equations for the entries of the matrix P_i. Therefore, aside from difficulties with high-order problems, calculation is straightforward.

Notice that the end result is not unreasonable; if the sequence $P_0, P_1, \ldots,$ approaches any limit at all, call it X, then $K_0, K_1, \ldots,$ approaches XGR^{-1}, and taking limits in (15.3-10) results in

$$X(F - GR^{-1}G'X) + (F' - XGR^{-1}G')X + XGR^{-1}G'X + Q = 0$$

or

$$XF + F'X - XGR^{-1}G'X + Q = 0.$$

To be sure, it is not clear that the sequence $P_0, P_1, \ldots,$ should approach a limit, or, indeed, that this limit should be positive definite (and therefore equal to $\bar{P}$). But, in fact, it can be shown [10] that

$$P_0 \geq P_1 \geq P_2 \geq \cdots \geq \bar{P}$$

from which it follows that the sequence does possess a positive definite limit. Reference [10] also shows that convergence of the sequence $P_0, P_1, \ldots,$ toward $\bar{P}$ is quadratic—i.e.,

$$\| P_{i+1} - \bar{P} \| \leq c \, \| P_i - \bar{P} \|^2$$

for some constant c and for all i. This guarantees rapidity of convergence when P_i is close to $\bar{P}$. A figure of 10 iterations is claimed as being typically required.

A second iterative procedure under development [11] starts by replacing the continuous time minimization problem with a discrete time minimization problem. The replacement is such that solution of the discrete time minimization problem leads to a solution, including knowledge of $\bar{P}$, of the continuous time problem. We now outline the procedure.

Applying the transformations

$$\begin{aligned}
\bar{\Phi} &= \tfrac{1}{2}(I - F')\bar{P}(I - F) \\
A &= (I - F)^{-1}(I + F) \\
B &= 2(I - F)^{-2}G \\
C &= R + G'(I - F')^{-1}Q(I - F)^{-1}G \\
D &= Q(I - F)^{-1}G
\end{aligned}$$

the quadratic matrix equation (15.3-1) may be written as

$$A'\bar{\Phi}A - \bar{\Phi} - [A'\bar{\Phi}B + D][C + B'\bar{\Phi}B]^{-1}[A'\bar{\Phi}B + D']' + Q = 0.$$

It has been shown that the existence of a unique nonnegative definite solution $\bar{P}$ of (15.3-1) implies the existence of a unique nonnegative definite solution Φ of the preceding equation. This solution may be determined by solving a discrete Riccati equation as follows:

$$\bar{\Phi} = \lim_{i \to \infty} \Phi_i$$

where

$$\begin{aligned}
\Phi_{i+1} &= A'\Phi_i A - [A'\Phi_i B + D][C + B'\Phi_i B]^{-1}[A'\Phi_i B + D]' + Q \\
\Phi_0 &= 0. \qquad (15.3\text{-}15)
\end{aligned}$$

It has been shown that the existence of a nonnegative definite solution of (15.3-1) implies convergence of this difference equation. Once $\bar{\Phi}$ is calculated (convergence typically requires 10 iterations), then $\bar{P}$ may be determined from $\bar{\Phi}$ by the transformation

$$\bar{P} = 2(I - F')^{-1}\bar{\Phi}(I - F)^{-1}.$$

We observe that at each iteration of the Riccati equation (15.3-15), the inverse of a matrix of dimension equal to the number of system inputs is required rather than the solution of a linear matrix equation of order equal to the state vector dimension, as in the previous method. Notice, also, that it is not required that an initial gain K be determined such that $F + GK'$ has negative eigenvalues. Detailed comparison with other methods is being investigated at the time of this manuscript's writing.

Finally, in this section we wish to show how a transient Riccati differ-

ential equation solution may be found from a steady state solution. In other words, we wish to find an expression for the solution $P(t)$ of

$$-\dot{P} = PF + F'P - PGR^{-1}G'P + Q \qquad P(0) = A \qquad (15.3\text{-}16)$$

given a positive definite $\bar{P}$ satisfying the limiting version of (15.3-16):

$$0 = \bar{P}F + F'\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q. \qquad (15.3\text{-}1)$$

The existence of an expression for $P(t)$ in terms of $\bar{P}$ means that the techniques of this section for the calculation of $\bar{P}$ now have extended capability; they may be added to those of the previous section for the calculation of $P(t)$.

With $\bar{P}$ the positive definite solution of (15.3-1) and

$$\bar{F} = F - GR^{-1}G'\bar{P} \qquad (15.3\text{-}17)$$

and with the assumption that $(A - \bar{P})^{-1}$ exists (which will certainly be the case if $A = 0$), we shall establish the following formula, valid for $t \leq 0$:

$$P(t) = \bar{P} + e^{-\bar{F}'t}[e^{-\bar{F}t}\bar{X}e^{-\bar{F}'t} + (A - \bar{P})^{-1} - \bar{X}]^{-1}e^{-\bar{F}t} \qquad (15.3\text{-}18)$$

where $\bar{X}$ is the unique negative definite solution of

$$\overline{F}\overline{X} + \overline{X}\overline{F}' - GR^{-1}G' = 0. \qquad (15.3\text{-}19)$$

The significance of $\bar{F}$ may be recalled from our earlier discussion of the optimal regulator. The control law for the optimal regulator is $u = -R^{-1}G'\bar{P}x$, and therefore the closed-loop regulator is described by $\dot{x} = \bar{F}x$. As we know, the closed-loop regulator is asymptotically stable; therefore, in (15.3-18), as t becomes more and more negative, the exponential terms decay.

To establish (15.3-18), we start with the two equations (15.3-16) and (15.3-1). By subtraction, it follows that

$$\begin{aligned}-\frac{d}{dt}(P - \bar{P}) &= (P - \bar{P})F + F'(P - \bar{P}) - (P - \bar{P})GR^{-1}G'(P - \bar{P}) \\ &\quad - PGR^{-1}G'\bar{P} - \bar{P}GR^{-1}G'P + 2\bar{P}GR^{-1}G'\bar{P} \\ &= (P - \bar{P})\bar{F} + \bar{F}'(P - \bar{P}) - (P - \bar{P})GR^{-1}G'(P - \bar{P}).\end{aligned}$$

Now assume that $(P - \bar{P})^{-1}$ exists for all t. Setting $X(t) = [P(t) - \bar{P}]^{-1}$, it follows that $\dot{X} = -(P - \bar{P})^{-1}[(d/dt)(P - \bar{P})](P - \bar{P})^{-1}$, and, therefore,

$$\dot{X} = \bar{F}X + X\bar{F}' - GR^{-1}G' \qquad (15.3\text{-}20)$$

with

$$X(0) = (A - \bar{P})^{-1}. \qquad (15.3\text{-}21)$$

{*Note:* Equation (15.3-20) always has a solution provided $X(0)$ exists. Consequently, the condition that $[P(t) - \bar{P}]^{-1}$ exists is that $(A - \bar{P})^{-1}$ exists.}

Now $\bar{X}$ satisfies the algebraic equation

$$0 = \bar{F}\bar{X} + \bar{X}\bar{F}' - GR^{-1}G'. \tag{15.3-19}$$

It so happens that with $\bar{F}$ possessing all its eigenvalues with negative real parts, this equation—really a set of linear equations for the entries of $\bar{X}$—always has a unique solution [21]. See also Appendix A.

Now subtract (15.3-19) from (15.3-20) to get

$$\frac{d}{dt}(X - \bar{X}) = \bar{F}(X - \bar{X}) + (X - \bar{X})\bar{F}'$$

with $X - \bar{X}$ having initial value $(A - \bar{P})^{-1} - \bar{X}$. It follows immediately from this equation that

$$X(t) - \bar{X} = e^{\bar{F}t}[(A - \bar{P})^{-1} - \bar{X}]e^{\bar{F}'t}.$$

Recalling now that $X(t)$ is $[P(t) - \bar{P}]^{-1}$, we deduce

$$\begin{aligned} P(t) &= \bar{P} + [\bar{X} + e^{\bar{F}t}(A - \bar{P})^{-1}e^{\bar{F}'t} - e^{\bar{F}t}\bar{X}e^{\bar{F}'t}]^{-1} \\ &= \bar{P} + e^{-\bar{F}'t}[e^{-\bar{F}t}\bar{X}e^{-\bar{F}'t} + (A - \bar{P})^{-1} - \bar{X}]^{-1}e^{-\bar{F}t} \end{aligned}$$

as required.

Problem 15.3-1. Using both methods discussed in this section, find the positive definite solution of

$$\bar{P}\begin{bmatrix} 0 & 0 \\ -1 & \sqrt{2} \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 0 & \sqrt{2} \end{bmatrix}\bar{P} - \bar{P}\begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \quad 0]\bar{P} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 0.$$

Problem 15.3-2. Consider the equation

$$\bar{P}F + F'\bar{P} - \bar{P}GR^{-1}G'\bar{P} + Q = 0.$$

Let

$$M = \begin{bmatrix} F & -GR^{-1}G' \\ -Q & -F' \end{bmatrix}$$

and let

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$$

be such that

$$TMT^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}.$$

Show that $\bar{P} = T_{21}T_{11}^{-1}$ is a solution of the quadratic matrix equation.

Problem 15.3-3. Using both methods of this section, find the positive definite solution of

$$\begin{bmatrix} 0 & f_{12} \\ f_{21} & 0 \end{bmatrix}\bar{P} + \bar{P}\begin{bmatrix} 0 & f_{21} \\ f_{12} & 0 \end{bmatrix} - \bar{P}\begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \quad 0]\bar{P} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = 0.$$

Assume $f_{12}f_{21} \neq 0$. Discuss also the cases $f_{12} = 0$ and $f_{21} = 0$.

Problem 15.3-4. Given the steady state solution

$$\bar{P} = \begin{bmatrix} 2-\sqrt{2} & 3-2\sqrt{2} \\ 3-2\sqrt{2} & 6-4\sqrt{2} \end{bmatrix}$$

of the equation

$$-\dot{P} = P\begin{bmatrix} 0 & 0 \\ 1 & -\sqrt{2} \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & -\sqrt{2} \end{bmatrix}P - P\begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \quad 0]P + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

find an explicit expression for $P(t)$ using the formula (15.3-18).

Problem 15.3-5. Find an expression for the solution of

$$-\dot{p} = 2pf - p^2g^2 + h^2$$

in terms of the positive solution of

$$2\bar{p}f - \bar{p}^2g^2 + h^2 = 0.$$

15.4 APPROXIMATE SOLUTIONS OF HIGH-ORDER RICCATI EQUATIONS VIA SINGULAR PERTURBATION

The computing time required to obtain a solution to high-order Riccati equations may be too long for some applications, and since the computing time increases at least according to the square of the system order, there is the temptation to neglect second-order effects, which increase the dimension of the state vector. This temptation is particularly strong when many solutions of a Riccati equation are required. However, experience shows that to carry out a design neglecting second-order effects altogether often leads to completely erroneous results.

Recent studies [22] and [23] have shown that the Riccati equation solutions obtained by neglecting the second-order effects, at least in two major areas of application, are not irrelevant to a problem solution. The addition of further relatively straightforward calculations enables an approximate solution to the high-order Riccati equation to be determined from the exact solution to the low-order equation or equations. For many design problems that occur in practice, this approach to obtaining a Riccati equation solution is quite adequate and considerable computing time can be saved.

We shall consider two cases that arise in control applications, where an nth order system (not necessarily time invariant)

$$\dot{x} = Fx + Gu \tag{15.4-1}$$

is to be controlled to minimize an index

$$V(x(t_0), u(\cdot), T) = \int_{t_0}^{T} (u'u + x'Qx)\, dt \tag{15.4-2}$$

where Q is nonnegative definite symmetric. (Without loss of generality, we

have assumed for simplicity that the usual $u'Ru$ term is simply $u'u$.) The Riccati equation of interest is

$$-\dot{P} = PF + F'P - PGG'P + Q \quad P(T, T) = 0. \tag{15.4-3}$$

The case of nominally "decoupled" systems. The first case considered is the case where the second-order effect consists of coupling between two nominally decoupled systems. That is, the equations for system (15.4-1) may be partitioned as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} F_1 & \epsilon F_{12} \\ \epsilon F_{21} & F_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} G_1 & \epsilon G_{12} \\ \epsilon G_{21} & G_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{15.4-5}$$

where ϵ is a scalar parameter, assumed small. If ϵ is set to zero, then (15.4-4) is the equation of two decoupled systems. The matrix Q in the performance index may be partitioned as

$$Q = \begin{bmatrix} Q_1 & \epsilon Q_{12} \\ \epsilon Q'_{12} & Q_2 \end{bmatrix} \tag{15.4-5}$$

and the associated Riccati equation solution P as

$$P = \begin{bmatrix} P_1 & P_{12} \\ P'_{12} & P_2 \end{bmatrix}. \tag{15.4-6}$$

We now adopt the notation $\tilde{P} = P|(\epsilon = 0)$, and $\tilde{P}_\epsilon = |(\partial P/\partial \epsilon)\,(\epsilon = 0)$. Differential equation theory yields that P is analytic in ϵ, and therefore P may be expanded as a Maclaurin series expansion as follows

$$P = \tilde{P} + \epsilon \tilde{P}_\epsilon + \text{Higher-order terms.} \tag{15.4-7}$$

Straightforward manipulations (the derivation of which is called for in Problem 15.4-1) yield that

$$\tilde{P} = \begin{bmatrix} \tilde{P}_1 & 0 \\ 0 & \tilde{P}_2 \end{bmatrix} \quad \tilde{P}_\epsilon = \begin{bmatrix} 0 & \tilde{P}_{12\epsilon} \\ \tilde{P}'_{12\epsilon} & 0 \end{bmatrix} \tag{15.4-8}$$

where

$$-\dot{\tilde{P}}_1 = \tilde{P}_1 F_1 + F'_1 \tilde{P}_1 - \tilde{P}_1 G_1 G'_1 \tilde{P}_1 + Q \qquad \tilde{P}_1(T, T) = 0 \tag{15.4-9}$$

$$-\dot{\tilde{P}}_2 = \tilde{P}_2 F_2 + F'_2 \tilde{P}_2 - \tilde{P}_2 G_2 G'_2 \tilde{P}_2 + Q_2 \qquad \tilde{P}_2(T, T) = 0 \tag{15.4-10}$$

$$-\dot{\tilde{P}}_{12\epsilon} = \tilde{P}_{12\epsilon}(F_2 - G_2 G'_2 \tilde{P}_2) + (F_1 - G_1 G'_1 \tilde{P}_1)'\tilde{P}_{12\epsilon} + Y$$

$$\tilde{P}_{12\epsilon}(T, T) = 0 \tag{15.4-11}$$

and

$$Y = \tilde{P}_1 F_{12} + F'_{21} \tilde{P}_2 - \tilde{P}_1 (G_1 G'_{21} + G_{12} G'_2) \tilde{P}_2 + Q_{12}. \tag{15.4-12}$$

When the nominally decoupled systems are assumed totally decoupled, the parameter ϵ is taken to be zero and P is approximated by $\tilde{P}$. The matrix $\tilde{P}$ is determined from two Riccati equations, (15.4-9) and (15.4-10), each of

lower order than (15.4-3). There are many situations where this approach results in unsatisfactory system performance—such as, e.g., where temperature, pressure, and flow control systems are designed independently despite the coupling that occurs between them in a chemical process [22].

For the case when coupling is small, the parameter ϵ is small, and an improved approximation for P is given from $P = \bar{P} + \epsilon\bar{P}_\epsilon$. To calculate $\bar{P}_\epsilon$ simply requires the solution of a linear equation (15.4-11) and (15.4-12). Higher-order terms in the Maclaurin series expansion may be taken into account without much further effort, as indicated in reference [22].

The corresponding time-invariant infinite-time problem is easy to solve. The limiting solutions of (15.4-9) and (15.4-10) exist if $[F_1, G_1]$ and $[F_2, G_2]$ are completely controllable, respectively. The limiting solution of (15.4-11) will exist if the eigenvalues of $F_1 - G_1G_1'\bar{P}_1$ and $F_2 - G_2G_2'\bar{P}_2$ all have negative real parts. As we know, sufficient conditions guaranteeing that such will be the case are that $[F_1, D_1]$ and $[F_2, D_2]$ are completely observable for any D_1 and D_2 with $D_1D_1' = Q_1$ and $D_2D_2' = Q_2$. These various conditions all have obvious physical significance.

In summary, then, an approximate method for solving Riccati equations associated with nominally decoupled systems has been given, which requires the solution of low-order Riccati equations and linear equations rather than the solution of a single high-order Riccati equation. For large-order systems, this represents a considerable saving in computer time. Examples and performance analysis details are given in [22].

The case of a nominally low-order system. The second solution we examine arises when the system under consideration is exactly described by high-order equations, but approximately described by low-order equations. The approximation is possible because some time constants are extremely fast, and their effect on the system's dynamics is almost negligible. In pole-zero terms for time-invariant systems, some of the poles of the system transfer function may be well in the left half-plane, with their residues in a partial fraction expansion being very small; the approximation would involve neglecting these poles.

We assume that the system equations (15.4-1) are the nominal or approximating low-order equations and that the performance index (15.4-2) is an index associated with this nominal low-order system. The Riccati equation (15.4-3) enables an optimal control to be obtained for the low-order system.

Now experience shows that when this control computed by neglecting the small time constants is applied to the actual physical system, unsatisfactory (perhaps unstable) performance often results [23]. To avoid this situation, the small time constants must be taken into account in some way. One way to do so is to avoid any approximations at all by using an exact high-order system description and by solving a high-order Riccati equation. However,

we shall steer a path intermediate between the exact analysis and an analysis involving total neglect of the fast time constants. This results in some computational saving while normally retaining satisfactory closed-loop performance.

The approach adopted for this problem is essentially the same as that adopted for the case of nominally coupled systems. Namely, we introduce a small parameter, now denoted by λ, to expand the solution of the Riccati equation associated with the high-order system as a Maclaurin series in λ, and we approximate this solution by taking into account the first few terms of the series expansion. The calculations involved in this approach require less computer time than if the high-order Riccati equation is solved directly.

As a first step, we express all the small time constants in terms of a parameter λ, where it is understood that λ has a known small value for the given system. For example, if a time constant τ has a value $\tau = 0.04$, it is written as $\tau = 0.4\lambda$ (say), where it is understood that λ has the value $\lambda = 0.1$ for the given system.

When the states associated with the small time constants are denoted by z, the high-order system description turns out to have the form

$$\begin{bmatrix} \dot{x} \\ \lambda\dot{z} \end{bmatrix} = \begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix}\begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}u. \tag{15.4-13}$$

In this equation, we stress again that λ is small compared to unity. Furthermore, the entries of the F_i and G_i, although possibly dependent on λ, have comparable magnitudes. In our derivation here, we shall assume that the F_i and G_i are independent of λ, as is often the case. Problem 15.4-4 asks for this restriction to be removed.

Taking λ equal to zero amounts to neglecting the fast time constants. Therefore, by setting $\lambda = 0$ in (15.4-13), we should be able to recover the low-order system equation (15.4-1). The mechanics of this operation proceed formally as follows. Assuming invertibility of F_4,

$$0 = F_3x + F_4z + G_2u$$

or

$$z = -F_4^{-1}(F_3x + G_2u) \tag{15.4-14}$$

and from (15.4-13), again,

$$\dot{x} = (F_1 - F_2F_4^{-1}F_3)x + (G_1 - F_2F_4^{-1}G_2)u. \tag{15.4-15}$$

The qualification that (15.4-13) and the pair (15.4-14) and (15.4-15) are only formally equivalent comes about in the following way. What is desired is that the solutions $[x(t, \lambda), z(t, \lambda)]$ of (15.4-13) should approach the solutions $[x(t), z(t)]$ of (15.4-14) and (15.4-15) as λ approaches zero. Reference [24] shows that this occurs for $t > t_0$ if and only if the eigenvalues of F_4 have nega-

tive real parts for all t. It may be thought that this is an unreasonable restriction to make, but, in actual fact, it is not, since it amounts to insisting (in the constant F_4 case) that all poles that we wish to neglect lie in the half-plane $\text{Re}[s] < 0$.

Equation (15.4-15) agrees with Eq. (15.4-1) if we put

$$F = F_1 - F_2F_4^{-1}F_3 \quad G = G_1 - F_2F_4^{-1}G_2. \tag{15.4-16}$$

Equation (15.4-16) may be regarded as a set of relations guaranteeing that (15.4-1) and (15.4-13) be *compatible representations* of the same system. They also provide a formal procedure for passing from the exact, high-order representation of the system to the nominal, approximate, low-order representation.

The performance index associated with the high-order system equations will be denoted by

$$V(x(t_0), z(t_0), u(\cdot), T) = \int_{t_0}^{T} (u'u + x'Qx)\, dt \tag{15.4-17}$$

where Q is nonnegative definite. Subsequently, we shall consider the time-invariant infinite-time problem. It is possible to take as a performance index for the high-order system an index where the loss function is of the form

$$u'u + [x' \quad z']Q\begin{bmatrix} x \\ z \end{bmatrix}$$

as shown in [25]. However, there is in general no need to include $z(\cdot)$ in the performance index, since the exponential rate of decay of $z(t)$ is assumed high a priori without any feedback control.

The solution of the Riccati equation associated with the high-order system (15.4-13) and the preceding index (15.4-17) will be denoted by Π. The matrix Π may be regarded as a function of λ as well as of t—i.e., each λ gives a different value of Π. In terms of a partition of Π as

$$\Pi = \begin{bmatrix} \Pi_1 & \lambda\Pi_2 \\ \lambda\Pi_2' & \lambda\Pi_3 \end{bmatrix} \tag{15.4-18}$$

and the use of the superscript tilde to denote evaluation at $\lambda = 0$ and the subscript λ to denote differentiation with respect to λ, our aims will be as follows:

1. To observe that the matrices Π_i are continuous at $\lambda = 0$, and that the $\tilde{\Pi}_i$ satisfy an equation derived from that satisfied by Π by formally setting $\lambda = 0$ in this latter equation.
2. To show by using the equations derived for the $\tilde{\Pi}_i$ that these matrices may be computed in terms of the solution P of the equation (15.4-3), associated with the low-order approximation to the control problem.

3. To observe that the matrices Π_i are continuously differentiable at $\lambda = 0$, and that the matrices $\tilde{\Pi}_{i\lambda}$ satisfy an equation derived from that satisfied by Π_λ by formally setting $\lambda = 0$ in this latter equation.
4. To observe that the matrices $\tilde{\Pi}_{i\lambda}$ may be derived by the solving of a linear matrix differential equation, and the carrying out of simple algebraic manipulations.

The conclusion is that Π may be approximated by

$$\Pi \doteq \begin{bmatrix} \tilde{\Pi}_1 + \lambda\tilde{\Pi}_{1\lambda} & \lambda(\tilde{\Pi}_2 + \lambda\tilde{\Pi}_{2\lambda}) \\ \lambda(\tilde{\Pi}'_2 + \lambda\tilde{\Pi}'_{2\lambda}) & \lambda(\tilde{\Pi}_3 + \lambda\tilde{\Pi}_{3\lambda}) \end{bmatrix} \tag{15.4-19}$$

provided λ is small, with the approximate value of Π far easier to compute than the exact value. The optimal feedback gain K for the high-order systems (15.4-13), normally given by

$$K = -\Pi \begin{bmatrix} G_1 \\ \dfrac{G_2}{\lambda} \end{bmatrix} \tag{15.4-20}$$

is instead approximated to first order in λ by

$$K \doteq -\begin{bmatrix} \tilde{\Pi}_1 G_1 + \lambda\tilde{\Pi}_{1\lambda} G_1 + \tilde{\Pi}_2 G_2 + \lambda\tilde{\Pi}_{2\lambda} G_2 \\ \lambda\tilde{\Pi}'_2 G_1 + \tilde{\Pi}_3 G_2 + \lambda\tilde{\Pi}_{3\lambda} G_2 \end{bmatrix}. \tag{15.4-21}$$

We now deal with each of the preceding aims 1 through 4, in turn. Using the high-order system equation (15.4-13), the loss function in (15.4-17), and the partitioning of Π in (15.4-18), we have

$$\begin{aligned} -\begin{bmatrix} \dot{\Pi}_1 & \lambda\dot{\Pi}_2 \\ \lambda\dot{\Pi}'_2 & \lambda\dot{\Pi}_3 \end{bmatrix} &= \begin{bmatrix} \Pi_1 F_1 + \Pi_2 F_3 & \Pi_1 F_2 + \Pi_2 F_4 \\ \lambda\Pi'_2 F_1 + \Pi_3 F_3 & \lambda\Pi'_2 F_2 + \Pi_3 F_4 \end{bmatrix} \\ &+ \begin{bmatrix} F'_1\Pi_1 + F'_3\Pi'_2 & \lambda F'_1\Pi_2 + F'_3\Pi_3 \\ F'_2\Pi_1 + F'_4\Pi'_2 & \lambda F'_2\Pi_2 + F'_4\Pi_3 \end{bmatrix} - \begin{bmatrix} \Pi_1 G_1 + \Pi_2 G_2 \\ \lambda\Pi'_2 G_1 + \Pi_3 G_2 \end{bmatrix} \\ &\times [G'_1\Pi_1 + G'_2\Pi'_2 \quad \lambda G'_1\Pi_2 + G'_2\Pi_3] + \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{15.4.22}$$

with $\Pi_i(T) = 0$ for $i = 1, 2, 3$.

Standard differential equation theory guarantees that the solution of this equation depends analytically on λ away from $\lambda = 0$. But because setting $\lambda = 0$ amounts to lowering the order of the differential equation—as is seen from examination of the left side of (15.4-22)—special techniques must be used to conclude even continuity at $\lambda = 0$. A theorem of [24], however, turns out to be applicable. Precisely because of the earlier condition imposed on F_4, it is possible to conclude that (1) the matrices Π_i are continuous at $\lambda = 0$ for all $t < T$, and that (2) the matrices $\tilde{\Pi}_i(t)$ satisfy the equations obtained by formally setting $\lambda = 0$ in (15.4-22):

$$\begin{bmatrix} -\frac{d}{dt}\tilde{\Pi}_1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \tilde{\Pi}_1 F_1 + \tilde{\Pi}_2 F_3 & \tilde{\Pi}_1 F_2 + \tilde{\Pi}_2 F_4 \\ \tilde{\Pi}_3 F_3 & \tilde{\Pi}_3 F_4 \end{bmatrix}$$
$$+ \begin{bmatrix} F_1'\tilde{\Pi}_1 + F_3'\tilde{\Pi}_2' & F_3'\tilde{\Pi}_3 \\ F_2'\tilde{\Pi}_1 + F_4'\tilde{\Pi}_2' & F_4'\tilde{\Pi}_3 \end{bmatrix} - \begin{bmatrix} \tilde{\Pi}_1 G_1 + \tilde{\Pi}_2 G_2 \\ \tilde{\Pi}_3 G_2 \end{bmatrix}$$
$$\times [G_1'\tilde{\Pi}_1 + G_2'\tilde{\Pi}_2' \quad G_2'\tilde{\Pi}_3] + \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix} \tag{15.4-23}$$

together with $\tilde{\Pi}_1(T) = 0$. This achieves aim 1.

We now set out to achieve aim 2—i.e., to use the equations (15.4-23) to derive formulas for the matrices $\tilde{\Pi}_i$ in terms of P, the solution of (15.4-3). First, we consider the 2-2 block of (15.4-23), which yields

$$\tilde{\Pi}_3 F_4 + F_4'\tilde{\Pi}_3 = \tilde{\Pi}_3 G_2 G_2' \tilde{\Pi}_3.$$

Now this equation can have a multiplicity of solutions, including, obviously, the zero solution. However, it is not difficult to show, using the eigenvalue constraint on F_4, that all solutions are nonpositive definite. Now for λ non-zero, the matrix Π is nonnegative definite for all t, and therefore so is the submatrix $\lambda\Pi_3$. Hence, Π_3 is nonnegative for all $\lambda \neq 0$, and since $\tilde{\Pi}_3$ is the continuous limit as λ approaches zero of $\tilde{\Pi}_3$, $\tilde{\Pi}_3$ must be nonnegative definite. Consequently, $\tilde{\Pi}_3$ must be simultaneously nonpositive and nonnegative definite. The only possibility is

$$\tilde{\Pi}_3 = 0. \tag{15.4-24}$$

Examination of the block in the 1-2 position of (15.4-23) yields that

$$\tilde{\Pi}_2 = -\tilde{\Pi}_1 F_2 F_4^{-1}. \tag{15.4-25}$$

When this expression for Π_2 is inserted into the block in the 1-1 position of (15.4-23), there results the following differential equation for $\tilde{\Pi}_1$:

$$-\frac{d}{dt}\tilde{\Pi}_1 = \tilde{\Pi}_1(F_1 - F_2F_4^{-1}F_3) + (F_1 - F_2F_4^{-1}F_3)'\tilde{\Pi}_1$$
$$- \tilde{\Pi}_1(G_1 - F_2F_4^{-1}G_2)(G_1 - F_2F_4^{-1}G_2)'\tilde{\Pi}_1 + Q.$$

On using the definitions (15.4-16) for F and G, we observe that this equation becomes identical with the differential equation (15.4-3) for P. The boundary conditions also match. Hence,

$$\tilde{\Pi}_1 = P. \tag{15.4-26}$$

Equations (15.4-24) through (15.4-26) fulfill the requirements set out in aim 2.

We pass on now to aim 3, requiring observation of the differentiability at $\lambda = 0$ of the matrices Π_i, and the derivation of equations satisfied by the matrices $\tilde{\Pi}_{i\lambda}$. As remarked earlier, solutions of the Riccati equation (15.4-22)

are analytic away from $\lambda = 0$. In particular, we can differentiate (15.4-22) to obtain

$$-\begin{bmatrix} \dot{\Pi}_{1\lambda} & \dot{\Pi}_2 + \lambda\dot{\Pi}_{2\lambda} \\ \dot{\Pi}'_2 + \lambda\dot{\Pi}'_{2\lambda} & \dot{\Pi}_3 + \lambda\dot{\Pi}_{3\lambda} \end{bmatrix}$$
$$= \begin{bmatrix} \Pi_{1\lambda}F_1 + \Pi_{2\lambda}F_3 & \Pi_{1\lambda}F_2 + \Pi_{2\lambda}F_4 \\ \Pi'_2F_1 + \lambda\Pi'_{2\lambda}F_1 + \Pi_{3\lambda}F_3 & \Pi'_2F_2 + \lambda\Pi'_{2\lambda}F_2 + \Pi_{3\lambda}F_4 \end{bmatrix}$$
$$+ \begin{bmatrix} F'_1\Pi_{1\lambda} + F'_3\Pi'_{2\lambda} & F'_1\Pi_2 + \lambda F'_1\Pi_{2\lambda} + F'_3\Pi_{3\lambda} \\ F'_2\Pi_{1\lambda} + F'_4\Pi'_{2\lambda} & F'_2\Pi_2 + \lambda F'_2\Pi_{2\lambda} + F'_4\Pi_{3\lambda} \end{bmatrix}$$
$$- \begin{bmatrix} \Pi_{1\lambda}G_1 + \Pi_{2\lambda}G_2 \\ \Pi'_2G_1 + \lambda\Pi'_{2\lambda}G_1 + \Pi_{3\lambda}G_2 \end{bmatrix}$$
$$\times [G'_1\Pi_1 + G'_2\Pi_2 \quad \lambda G'_1\Pi_2 + G'_2\Pi_3] - \begin{bmatrix} \Pi_1G_1 + \Pi_2G_2 \\ \lambda\Pi'_2G_1 + \Pi_3G_2 \end{bmatrix}$$
$$\times [G'_1\Pi_{1\lambda} + G'_2\Pi'_{2\lambda} \quad G'_1\Pi_2 + \lambda G'_1\Pi_{2\lambda} + G'_2\Pi_{3\lambda}] \tag{15.4-27}$$

with $\Pi_{i\lambda}(T) = 0$ for $i = 1, 2, 3$.

Once again, the results of [24] combine with the properties of F_4 to yield the conclusion that the matrices $\Pi_{i\lambda}$ are continuous at $\lambda = 0$, and that equations satisfied by the matrices $\tilde{\Pi}_{i\lambda}$ may be obtained by formally setting $\lambda = 0$ in (15.4-27). These equations are [with the matrix X defined in (15.4-31)]

$$-\frac{d}{dt}\tilde{\Pi}_{1\lambda} = \tilde{\Pi}_{1\lambda}[F_1 - G_1(G'_1P + G'_2\tilde{\Pi}'_2)]$$
$$+ [F_1 - G_1(G'_1P + G'_2\tilde{\Pi}'_2)]'\tilde{\Pi}_{1\lambda} + X \tag{15.4-28}$$

$$\tilde{\Pi}_{2\lambda} = \Big[\frac{d}{dt}\tilde{\Pi}_2 - \tilde{\Pi}_{1\lambda}F_2 - F'_1\tilde{\Pi}_2$$
$$+ (PG_1 + \tilde{\Pi}_2G_2)(G'_1\tilde{\Pi}_2 + G'_2\tilde{\Pi}_{3\lambda})\Big]F_4^{-1} \tag{15.4-29}$$

and

$$0 = \tilde{\Pi}_{3\lambda}F_4 + F'_4\tilde{\Pi}_{3\lambda} + (F'_2\tilde{\Pi}_2 + \tilde{\Pi}'_2F_2) \tag{15.4-30}$$

with the initial condition $\tilde{\Pi}_{1\lambda}(T) = 0$. In equation (15.4-28), the quantity X is given by

$$X = \tilde{\Pi}_{2\lambda}F_3 + F'_3\Pi'_{2\lambda} - (PG_1 + \tilde{\Pi}_2G_2)G'_2\tilde{\Pi}'_{2\lambda}$$
$$- \tilde{\Pi}_{2\lambda}G_2(G'_1P + G'_2\tilde{\Pi}'_2) \tag{15.4-31}$$

The derivation of these equations means that aim 3 is achieved. Achievement of aim 4 is immediate: Eq. (15.4-30) is a readily solvable linear matrix equation yielding $\tilde{\Pi}_{3\lambda}$. This then allows elimination of $\tilde{\Pi}_{2\lambda}$ from (15.4-28) by using (15.4-29); then equation (15.4-28) becomes a linear matrix differential equation in $\tilde{\Pi}_{1\lambda}$, and is readily solved.

The time-invariant problem is also straightforward to solve. The matrices F_i, G_i, and Q are, of course, constant, and $T = +\infty$. One then seeks limiting

solutions of the Riccati equation for P and a limiting solution of the linear differential equation (15.4-28) for $\tilde{\Pi}_{1\lambda}$. Since this equation in its limiting form is a linear matrix equation, it is probably easier to determine its solution by solving the algebraic equation obtained by setting $(d/dt)\tilde{\Pi}_{1\lambda}$ equal to zero, than by obtaining the limiting differential equation solution. Let us now summarize the calculations required.

1. We start with a high-order system as in (15.4-13), with the F_i independent of λ. {If the F_i depend continuously on λ at $\lambda = 0$, we must use more complex formulas the development of which is called for in Problem 15.4-4, or which may be obtained from [23].} We aim to minimize the performance index (15.4-17).
2. We compute the equations of the nominal low-order system (15.4-1) using (15.4-16) and the matrix P of Eq. (15.4-3).
3. We compute the matrices $\tilde{\Pi}_i$ via (15.4-24) through (15.4-26).
4. We compute the matrices $\tilde{\Pi}_{i\lambda}$ via (15.4-28) through (15.4-31).
5. The matrix Π is approximated using the approximate values for the matrices $\Pi_i(\lambda)$ in (15.4-19); the approximate value of Π is used to determine an approximate control law in (15.4-21).

Reference [23] includes some numerical examples, which compare exact designs with approximate designs obtained by taking $\lambda = 0$, and by using the matrices $\tilde{\Pi}_i$ and $\tilde{\Pi}_{i\lambda}$ as discussed herein. Better approximation to the exact design could no doubt be obtained in these examples, and indeed in all situations, by computing $\tilde{\Pi}_{i\lambda\lambda}$ and higher derivatives. The computational task, however, will probably approach that associated with an exact design and negate the utility of such an approach.

We now present a simple example illustrating use of the preceding ideas. Although the example does not demonstrate large savings in computation (because in this case the "high-order" system is not really high order), it does indicate the sequence of calculations required.

The system we consider, shown in Fig. 15.4-1, is

$$\begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -10 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} 0 \\ 5 \end{bmatrix} u$$

which we choose to write as

$$\begin{bmatrix} \dot{x} \\ \lambda\dot{z} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix} u$$

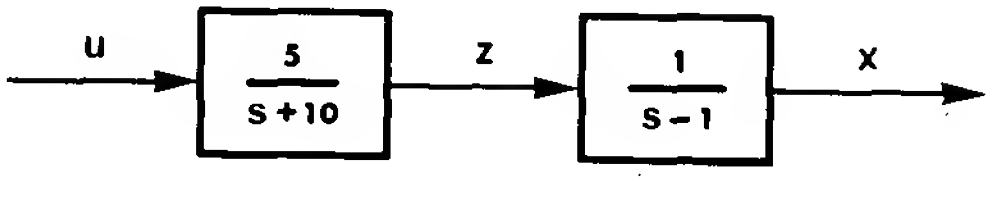


Fig. 15.4-1 The high-order system.

where $\lambda = 0.1$. The performance index we shall take to be

$$V = \int_{t_0}^{\infty} (x^2 + u^2)\, dt.$$

The first requirement is to obtain the associated nominally low-order system. Straightforward application of (15.4-16) yields this to be

$$\dot{x} = x + \tfrac{1}{2}u.$$

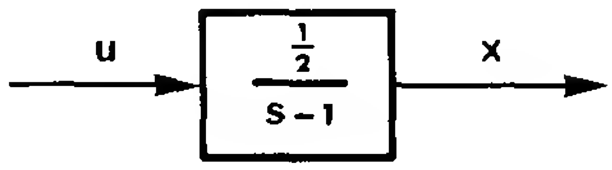


Fig. 15.4-2 The approximate low-order system.

Next, we solve the optimization problem associated with this low-order system. The optimal performance index is Px^2, where P is the positive definite root of

$$2P - \tfrac{1}{4}P^2 + 1 = 0$$

i.e.,

$$P = 4 + 2\sqrt{5}.$$

Applying (15.4-24) through (15.4-26), we obtain

$$\tilde{\Pi}_1 = 4 + 2\sqrt{5}, \quad \tilde{\Pi}_2 = 4 + 2\sqrt{5}, \quad \tilde{\Pi}_3 = 0.$$

Equations (15.4-28) through (15.4-31) are now used to obtain the matrices $\tilde{\Pi}_{i\lambda}$. First, from Eq. (15.4-30), we have

$$\tilde{\Pi}_{3\lambda} = (4 + 2\sqrt{5}).$$

Equation (15.4-29) yields

$$\tilde{\Pi}_{2\lambda} = \tilde{\Pi}_{1\lambda} + (5 + 2\sqrt{5}).$$

Substituting into the steady state version of (15.4-28), we obtain

$$\tilde{\Pi}_{1\lambda} = -9 - 4\sqrt{5}.$$

Thus,

$$\tilde{\Pi}_{2\lambda} = -4 - 2\sqrt{5}.$$

From Eq. (15.4-21), the approximate feedback gain vector is

$$K = -\begin{bmatrix} (2 + \sqrt{5}) - \lambda(2 + \sqrt{5}) \\ \lambda(2 + \sqrt{5}) \end{bmatrix}.$$

The difference is clear between taking $\lambda = 0$ (corresponding to total neglect of the dynamics associated with z) and $\lambda = 0.1$, (corresponding to an approximation consideration of the dynamics associated with z). Problem 15.4-5 asks for a comparison with the exact feedback gain.

Problem 15.4-1. Using Eqs. (15.4-3) through (15.4-6) of the text, derive Eqs. (15.4-9) through (15.4-12) for $\tilde{P}$ and $\tilde{P}_\epsilon$.

Problem 15.4-2. Derive recursive formulas for

$$\left.\frac{\partial^{2i} P}{\partial \epsilon^{2i}}\right| (\epsilon = 0)$$

and

$$\left.\frac{\partial^{2i+1} P}{\partial \epsilon^{2i+1}}\right| (\epsilon = 0)$$

given Eqs. (15.4-3) through (15.4-6).

Problem 15.4-3. Given a plant with a transfer function

$$W(s) = \frac{3}{0.1s + 1} \cdot \frac{3}{0.04s + 1} \cdot \frac{6}{0.07s + 1} \cdot \frac{32}{2s + 1} \cdot \frac{25}{5s + 1}$$

express the state-space equation for the system in the form of (15.4-13).

Problem 15.4-4. Assume that the matrices F_i in Eq. (15.4-13) depend continuously on λ in the vicinity of $\lambda = 0$. Indicate the variations necessary in the subsequent derivation of the formulas for $\tilde{\Pi}_{i\lambda}$.

Problem 15.4-5. Compare the approximate feedback gain and performance index obtained in the worked example with the exact feedback gain and performance index.

Problem 15.4-6. Can you suggest how the results of this section can be applied to optimal filtering problems?

REFERENCES

[1] Kalman, R. E., "Contributions to the Theory of Optimal Control," *Bol. Soc. Matem. Mex.*, 1960, pp. 102–119.

[2] Kalman, R. E., and T. S. Englar, "A User's Manual for ASP C," *RIAS Rep. NASA*, 1965, Baltimore, Md.

[3] Reid, W. T., "A Matrix Differential Equation of the Riccati Type," *Am. J. Math.*, Vol. 68, 1946, pp. 237–246.

[4] Levin, J. J., "On the Matrix Riccati Equation," *Proc. Am. Math. Soc.*, Vol. 10, 1959, pp. 519–524.

[5] O'Donnell, J. J., "Asymptotic Solution of the Matrix Riccati Equation of Optimal Control," *Proc. Fourth Ann. Allerton Conf. Circuit System Theory*, October 1966, Urbana, Ill., pp. 577–586.

[6] Vaughan, D. R., "A Negative Exponential Solution to the Matrix Riccati Equation," *IEEE Trans. Auto. Control*, Vol. AC-14, No. 1, February 1969, pp. 72–75.

[7] Potter, J. E., "Matrix Quadratic Solutions," *SIAM J. Applied Math.*, Vol. 14, No. 3, May 1966, pp. 496–501.

[8] Macfarlane, A. G. J., "An Eigenvector Solution of the Optimal Linear Regulator," *J. Electron. Control*, Vol. 14, No. 6, June 1963, pp. 643–654.

[9] Bucy, R. S., and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance*, Interscience Publishers, New York, 1968.

[10] Kleinman, D. L., "On an Iterative Technique for Riccati Equation Computation," *IEEE Trans. Auto. Control*, Vol. AC-13, No. 1, February 1968, pp. 114–115.

[11] Hitz, K. L., and B. D. O. Anderson, "The Relation between Continuous and Discrete Time Quadratic Minimization Problems," *Proc. Hawaii Conf. Systems Sci.*, 1969, pp. 847–850.

[12] Hendricks, T. C., and G. W. Haynes, The "GASP" Computer Program, *Proc. Second Asilomar Conf. Circuits Systems*, 1968, Monterey, Calif.

[13] Wonham, W. M., "On Pole Assignment in Multi-Input Controllable Linear Systems," *IEEE Trans. Auto. Control*, Vol. AC-12, No. 6, December 1967, pp. 660–665.

[14] Luenberger, D. G., "Canonical Forms for Linear Multivariable Systems," *IEEE Trans. Auto. Control*, Vol. AC-12, No. 3, June 1967, pp. 290–293.

[15] Popov, V. M., "Hyperstability and Optimality of Automatic Systems with Several Control Functions," *Rev. Roumaine Sci. Tech.—Electrotechnol. Energ.*, Vol. 9, No. 4, 1964, pp. 629–690.

[16] Anderson, B. D. O., and D. G. Luenberger, "Design of Multivariable Feedback Systems," *Proc. IEE*, Vol. 114, No. 3, March 1967, pp. 395–399.

[17] Luenberger, D. G., "Observers for Multivariable Systems," *IEEE Trans. Auto. Control*, Vol. AC-11, No. 2, April 1966, pp. 190–197.

[18] Gopinath, B., and B. O. Lange, "On the Identification and Control of Linear Systems," *Tech. Rep. SUDAAR No. 351*, June 1968, Dept. of Aeronautics and Astronautics, Stanford Univ., Stanford., Calif.

[19] Rosenbrock, H. H., "The Assignment of Closed Loop Poles," *Control Systems Centre Rep. No. 20*, October 1967, Control Systems Centre, University of Manchester, England.

[20] Singer, R. A., "The Design and Synthesis of Linear Multivariable Systems with Application to State Estimation," *Tech. Rep. 6302-8*, June 1968, Stanford Electronics Laboratories, Stanford, Calif.

[21] Gantmacher, F. R., *The Theory of Matrices*, Chelsea Publishing Co., New York, 1959.

[22] Kokotović, P. V., W. R. Perkins, J. B. Cruz, and G. D'Ans, "ϵ—Coupling Method for Near-Optimum Design of Large Scale Linear Systems," *Proc. IEE*, Vol. 116, No. 5, May 1969, pp. 889–892.

[23] Sannuti, P., and P. V. Kokotović, "Near-Optimum Design of Linear Systems by a Singular Perturbation Method," *IEEE Trans. Auto. Control*, Vol. AC–14, No. 1, February 1969, pp. 15–22.

[24] Vasileva, A. B., "Asymptotic Behaviour of Solutions to Certain Problems Involving Nonlinear Differential Equations Containing a Small Parameter Multiplying the Highest Derivative," *Russ. Math. Surv.*, Vol. 18, No. 3, 1963, pp. 13–81.

[25] Kokotović, P. V., and P. Sannuti, "Singular Perturbation Method for Reducing the Model Order in Optimal Control Design," *IEEE Trans. Auto. Control*, Vol. AC–13, No. 4, August 1968, pp. 377–384.

APPENDIX A

BRIEF REVIEW OF SOME RESULTS OF MATRIX THEORY

The purpose of this appendix is to provide a rapid statement of those particular results of matrix theory used in this book. For more extensive treatments standard text books—e.g., [1] and [2]—should be consulted. The latter is a particularly useful source of computational techniques.

1. Matrices and vectors. An $m \times n$ *matrix* A consists of a collection of mn quantities† a_{ij} $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ written in an array of m rows and n columns:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Sometimes, one simply writes

$$A = (a_{ij}).$$

The quantity a_{ij} is an *entry* (the $(i\text{-}j)$th entry, in fact) of A.

†The a_{ij} will be assumed real in most of our discussions.

An *m vector*, or, more fully, a *column m vector*, is a matrix with 1 column and m rows; thus,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{bmatrix}$$

defines x as column m vector, whose ith entry is the quantity x_i. A row n vector is a matrix with 1 row and n columns.

2. Addition, subtraction, and multiplication by a scalar. Two matrices A and B with the *same number of rows and also the same number of columns* may be added, subtracted, or individually multiplied by a scalar. With k_1, k_2, scalar, the matrix

$$C = k_1A + k_2B$$

is defined by

$$c_{ij} = k_1a_{ij} + k_2b_{ij}.$$

Thus, to add two matrices, one simply adds corresponding entries; to subtract two matrices, one simply subtracts corresponding entries; etc. Of course, addition is commutative—i.e.,

$$A + B = B + A.$$

3. Multiplication of matrices. Consider two matrices A and B, with A an $m \times p$ matrix and B a $p \times n$ matrix. Thus, the number of columns of A equals the number of rows of B. The product AB is an $m \times n$ matrix defined by

$$C = AB$$

with

$$c_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj}.$$

Notice that C has the same number of rows as A, and the same number of columns as B.

The product of three (or more) matrices can be defined by

$$D = ABC = (AB)C = A(BC).$$

In other words, multiplication is associative. However, multiplication is not commutative—i.e. it is *not* in general true that

$$AB = BA.$$

In fact, although AB can be formed, the product BA may not be capable of being formed.

For any integer p, the $p \times p$ matrix

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

possessing p rows and columns is termed the *identity matrix of order p*. It has the property that with A any $m \times p$ matrix,

$$AI = A.$$

Likewise, the identity matrix of order m has the property that

$$IA = A.$$

Any matrix consisting entirely of entries that are zero is termed *the zero matrix*. Its product with any matrix produces the zero matrix, whereas if it is added to any matrix, it leaves that matrix unaltered.

Suppose A and B are both $n \times n$ matrices (A and B are then termed square matrices). Then AB is square. It can be proved then that

$$|AB| = |A|\,|B|$$

where $|A|$ is the determinant of A.

[The definition of the determinant of a square matrix is standard. One way of recursively defining $|A|$ for A an $n \times n$ matrix is to expand A by its first row, thus

$$|A| = a_{11} \begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n2} & \cdots & a_{nn} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{33} & a_{34} & \cdots & a_{3n} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ a_{n1} & a_{n3} & a_{n4} & \cdots & a_{nn} \end{vmatrix}$$

$$+ a_{13} \begin{vmatrix} a_{21} & a_{22} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{34} & \cdots & a_{3n} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n4} & \cdots & a_{nn} \end{vmatrix} - \cdots$$

This expresses $|A|$ in terms of determinants of $(n-1) \times (n-1)$ matrices. In turn, these determinants may be expressed using determinants of $(n-2) \times (n-2)$ matrices, etc.]

4. Direct sum of two matrices. Let A be an $n \times n$ matrix and B an $m \times m$ matrix. The *direct* sum of A and B, written $A \dot{+} B$, is the $(n + m) \times (n + m)$ matrix

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

5. Transposition. Suppose A is an $m \times n$ matrix. The *transpose* of A, written A', is an $n \times m$ matrix defined by

$$B = A'$$

where

$$b_{ij} = a_{ji}.$$

Thus, if

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 5 \end{bmatrix},$$

$$A' = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 5 \end{bmatrix}.$$

It is easy to establish the important result

$$(AB)' = B'A',$$

which extends to

$$(ABC)' = C'B'A'$$

and so on. Also, trivially, one has

$$(A + B)' = A' + B'$$

6. Singularity and nonsingularity. Suppose A is an $n \times n$ matrix. Then A is said to be *singular* if $|A|$ is zero. Otherwise, A is termed nonsingular.

7. Rank of a matrix. Let A be an $m \times n$ matrix. The rank of A is a positive integer q such that some $q \times q$ submatrix of A, formed by deleting $(m - q)$ rows and $(n - q)$ columns, is nonsingular, whereas no $(q + 1) \times (q + 1)$ submatrix is nonsingular. For example, consider

$$A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 0 \end{bmatrix}.$$

The maximum size square submatrix that can be formed is 2×2. Therefore, a priori, rank $A \leq 2$. Now the possible 2×2 submatrices are

$$\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix} \quad \begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 \\ 3 & 0 \end{bmatrix}.$$

These all have zero determinant. Therefore, rank $A < 2$. Of the 1×1 submatrices, two have zero determinant but six do not. Therefore, rank $A = 1$.

The rank of A is also the maximum number of linearly independent rows of A and the maximum number of linearly independent columns of A. In the example, the second row equals the first row. Furthermore, the second, third, and fourth columns are linear multiples of the first.

It can be shown that

$$\operatorname{rank}(AB) \leq \min[\operatorname{rank} A, \operatorname{rank} B]$$

If rank A is equal to the number of columns *or* the number of rows of A, A is often said to have *full rank*. If A is $n \times n$, the statement rank $A = n$ is equivalent to the statement A is nonsingular. If, for an arbitrary matrix A, rank $A = 0$, then A is the zero matrix.

8. Range space and null space of a matrix. Let A be an $m \times n$ matrix. The range space of A, written $\mathscr{R}[A]$, is the set of all vectors Ax, where x ranges over the set of all n vectors. The range space has dimension equal to the rank of A—i.e., the maximal number of linearly independent vectors in $\mathscr{R}[A]$ is rank A. The nullspace of A, written $\mathscr{N}[A]$, is the set of vectors y for which $Ay = 0$.

An easily proved property is that $\mathscr{R}[A']$ and $\mathscr{N}[A]$ are orthogonal—i.e., if $y_1 = A'x$ for some x, and if y_2 is such that $Ay_2 = 0$, then $y_1'y_2 = 0$.

9. Inverse of a square nonsingular matrix. Let A be a square matrix. If, but only if, A is nonsingular, there exists a unique matrix, call it B, termed the *inverse of A*, with the properties

$$BA = AB = I.$$

The inverse of A is generally written A^{-1}. There are many computational procedures for passing from a prescribed A to its inverse A^{-1}. A formula is, in fact, available for the entries of $B = A^{-1}$, obtainable as follows.

Define the cofactor of the $i-j$ entry of A as $(-1)^{i+j}$ times the determinant of the matrix obtained by deleting from A the ith row and jth column, i.e., the row and column containing a_{ij}. Then,

$$b_{ij} = \frac{1}{|A|} \times \text{cofactor of } a_{ji}.$$

It easily follows that

$$(A^{-1})' = (A')^{-1}.$$

If A_1 and A_2 are two $n \times n$ nonsingular matrices, it can be shown that

$$(A_1A_2)^{-1} = A_2^{-1}A_1^{-1}.$$

10. Powers of a square matrix. For positive m, A^m for a square matrix A is defined as $AA \cdots A$, there being m terms in the product. For negative m, let $m = -n$, where n is positive; then $A^m = (A^{-1})^n$. It follows that $A^pA^q = A^{p+q}$ for any integers p and q, positive or negative, and likewise that $(A^p)^q = A^{pq}$.

A polynomial in A is a matrix $p(A) = \sum_{i=0}^{r} a_i A^i$ where the a_i are scalars. Any two polynomials in the same matrix commute—i.e., $p(A)q(A) = q(A)p(A)$, where p and q are polynomials. It follows that $p(A)q^{-1}(A) = q^{-1}(A)p(A)$, and that such *rational functions* of A also commute.

11. Exponential of a square matrix. Let A be a square matrix. Then it can be shown that the series

$$I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

converges, in the sense that the i-j entry of the partial sums of the series converges for all i and j. The sum is defined as e^A. It follows that

$$e^{At} = I + At + \frac{1}{2!}A^2t^2 + \cdots$$

Other properties are: $p(A)e^{At} = e^{At}p(A)$ for any polynomial A, and $e^{-At} = [e^{At}]^{-1}$.

12. Differentiation and integration. Suppose A is a function of a scalar variable t, in the sense that each entry of A is a function of t. Then

$$\frac{dA}{dt} = \left(\frac{da_{ij}}{dt}\right).$$

It follows that

$$\frac{d}{dt}(AB) = \frac{dA}{dt}B + A\frac{dB}{dt}.$$

Also, from the definition of e^{At}, one has

$$\frac{d}{dt}(e^{At}) = Ae^{At} = e^{At}A.$$

The integral of a matrix is defined in a straightforward way as

$$\int A\,dt = \left(\int a_{ij}\,dt\right).$$

Suppose ϕ is a scalar function of a vector x. Then

$$\frac{d\phi}{dx} = \text{a vector whose } i\text{th entry is } \frac{\partial\phi}{\partial x_i}.$$

Suppose ϕ is a scalar function of a matrix A. Then,

$$\frac{d\phi}{dA} = \text{a matrix whose } i\text{-}j \text{ entry is } \frac{\partial\phi}{\partial a_{ij}}.$$

Suppose z is a vector function of a vector x. Then,

$$\frac{dz}{dx} = \text{A matrix whose } i\text{-}j \text{ entry is } \frac{\partial z_i}{\partial x_j}.$$

13. Eigenvalues and eigenvectors of a square matrix. Let A be an $n \times n$ matrix. Construct the polynomial $|sI - A|$. This is termed the *characteristic polynomial* of A; the zeros of this polynomial are the *eigenvalues* of A. If λ_i is an eigenvalue of A, there always exists at least one vector x satisfying the equation

$$Ax = \lambda_i x.$$

The vector x is termed an eigenvector of the matrix A. If λ_i is not a repeated eigenvalue—i.e., if it is a simple zero of the characteristic polynomial, to within a scalar multiple x is unique. If not, there *may* be more than one eigenvector associated with λ_i. If λ_i is real, the entries of x are real, whereas if λ_i is complex, the entries of x are complex.

If A has zero entries everywhere off the main diagonal—i.e., if $a_{ij} = 0$ for all i, j, with $i \neq j$, then A is termed *diagonal*. (*Note:* Zero entries are still permitted on the main diagonal.) It follows trivially from the definition of an eigenvalue that the diagonal entries of the diagonal A are precisely the eigenvalues of A.

It is also true that for a general A,

$$|A| = \prod_{i=1}^{n} \lambda_i.$$

If A is singular, A possesses at least one zero eigenvalue.

The eigenvalues of a rational function $r(A)$ of A are the numbers $r(\lambda_i)$, where λ_i are the eigenvalues of A. The eigenvalues of e^{At} are $e^{\lambda_i t}$.

14. Trace of a square matrix A. Let A be $n \times n$. Then the trace of A, written $\operatorname{tr}[A]$, is defined as

$$\operatorname{tr}[A] = \sum_{i=1}^{n} a_{ii}.$$

An important property is that

$$\operatorname{tr}[A] = \sum_{i=1}^{n} \lambda_i$$

where the λ_i are eigenvalues of A. Other properties are

$$\operatorname{tr}[A + B] = \operatorname{tr}[B + A] = \operatorname{tr}[A] + \operatorname{tr}[B]$$

and, assuming the multiplications can be performed to yield square product matrices,

$$\operatorname{tr}[AB] = \operatorname{tr}[B'A'] = \operatorname{tr}[BA] = \operatorname{tr}[A'B']$$

$$\operatorname{tr}[A'A] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2.$$

15. Orthogonal, symmetric, and skew-symmetric matrices, and their eigenvalue properties. If a square A is such that $AA' = I$, and thus $A'A = I$, A is termed *orthogonal.* The eigenvalues of A then have a magnitude of unity. If $A = A'$, A is termed *symmetric*, and the eigenvalues of A are all real. Moreover, if x_1 is an eigenvector associated with λ_1, x_2 with λ_2, and if $\lambda_1 \neq \lambda_2$, then $x_1'x_2 = 0$. The vectors x_1 and x_2 are termed *orthogonal.* (*Note:* Distinguish between an orthogonal matrix and an orthogonal pair of vectors.) If $A = -A'$, A is termed *skew*, or *skew symmetric*, and the eigenvalues of A are pure imaginary.

16. The Cayley-Hamilton theorem. Let A be a square matrix, and let $|sI - A| = s^n + \alpha_1 s^{n-1} + \cdots + \alpha_n$; then,

$$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_n I = 0.$$

The Cayley–Hamilton theorem is often stated, rather ambiguously, as "the matrix A satisfies its characteristic polynomial."

From the Cayley–Hamilton theorem, it follows that A^m for any $m > n$ and e^A are expressible as a linear combination of $I, A, \ldots, A^{n-1}$.

17. Similar matrices and diagonalizability. Let A and B be $n \times n$ matrices. If there exists a nonsingular $n \times n$ matrix T such that $B = T^{-1}AT$, the matrices A and B are termed *similar.* Similarity is an equivalence relation. Thus:

1. A is similar to A.
2. If A is similar to B, then B is similar to A.
3. If A is similar to B and B is similar to C, then A is similar to C.

Similar matrices have the same eigenvalues. This may be verified by observing that

$$sI - B = T^{-1}sIT - T^{-1}AT = T^{-1}(sI - A)T.$$

Therefore,

$$|sI - B| = |T^{-1}||sI - A||T| = |sI - A||T^{-1}||T|.$$

But $T^{-1}T = I$ so that $|T^{-1}||T| = 1$. The result is then immediate.

If, for a given A, a matrix T can be formed such that

$$\Lambda = T^{-1}AT$$

is diagonal, then A is termed *diagonalizable*, the diagonal entries of Λ are eigenvalues of A, and the columns of T turn out to be eigenvalues of A.

Not all square matrices are diagonalizable, but matrices that have no repeated eigenvalues are diagonalizable, as are orthogonal, symmetric, and skew-symmetric matrices.

18. Jordan form. Not all square matrices are diagonalizable. But it is always possible to get very close to diagonal matrix via a *similarity trans-*

formation. In fact, there always exists a matrix T such that

$$T^{-1}AT = \begin{bmatrix} \lambda_1 & 1 & & & & & & & & & \\ & \lambda_1 & & & & & & & & & \\ & & \lambda_1 & & & & & & & & \\ & & & \lambda_2 & 1 & & & & & & \\ & & & & \lambda_2 & & & & & & \\ & & & & & \lambda_3 & 1 & & & & \\ & & & & & & \lambda_3 & 1 & & & \\ & & & & & & & \lambda_3 & & & \\ & & & & & & & & \cdot & & \\ & & & & & & & & & \cdot & \\ & & & & & & & & & & \cdot \end{bmatrix}$$

or something similar. Here, all blank entries are zero, the eigenvalues of A occur on the main diagonal, and there *may or may not* be entries of 1 above and to the right of repeated eigenvalues—i.e., on the superdiagonal. For any A, the distribution of 1's and 0's on the superdiagonal is fixed, but different A yield different distributions. The preceding almost-diagonal matrix is called the *Jordan canonical form* of A. The *Jordan blocks* of A are the matrices

$$\begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix} \quad [\lambda_1] \quad \begin{bmatrix} \lambda_2 & 1 \\ 0 & \lambda_2 \end{bmatrix} \quad \begin{bmatrix} \lambda_3 & 1 & 0 \\ 0 & \lambda_3 & 1 \\ 0 & 0 & \lambda_3 \end{bmatrix} \quad \text{etc.}$$

19. Positive and nonnegative definite matrices. Suppose A is $n \times n$ and symmetric. Then A is termed positive definite, if for all nonzero vectors x the scalar quantity $x'Ax$ is positive. Also, A is termed nonnegative definite if $x'Ax$ is simply nonnegative for all nonzero x. Negative definite and nonpositive definite are defined similarly. The quantity $x'Ax$ is termed a *quadratic form*, because when written as

$$x'Ax = \sum_{i,j=1}^{n} a_{ij}x_ix_j$$

it is quadratic in the entries x_i of x.

There are simple tests for positive and nonnegative definiteness. For A to be positive definite, all *leading minors* must be positive—i.e.,

$$a_{11} > 0 \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} > 0 \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} > 0 \quad \text{etc.}$$

For A to be nonnegative definite, all minors whose diagonal entries are diagonal entries of A must be nonnegative. That is, for a 3×3 matrix A,

$$a_{11}, a_{22}, a_{33} \geq 0 \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{13} \\ a_{13} & a_{33} \end{vmatrix}, \begin{vmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{vmatrix} \geq 0$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} \geq 0.$$

A symmetric A is positive definite if and only if its eigenvalues are positive, and nonnegative definite if and only if its eigenvalues are nonnegative.

If D is an $n \times m$ matrix, then $A = DD'$ is nonnegative definite, and positive definite if and only if D has rank n. An easy way to see this is to define a vector y by $y = D'x$. Then $x'Ax = xDD'x = y'y = \sum y_i^2 \geq 0$. The inequality becomes an equality if and only if $y = 0$ or $D'x = 0$, which is impossible for nonzero x if D has rank n.

If A is nonnegative definite, there exists a matrix B that is a *symmetric square root* of A; it is also nonnegative definite. It has the property that

$$B^2 = A$$

and is often denoted by $A^{1/2}$. If A is positive definite, so is $A^{1/2}$, and $A^{1/2}$ is then unique.

If A and B are nonnegative definite, so is $A + B$, and if one is positive definite, so is $A + B$. If A is nonnegative definite and $n \times n$, and B is $m \times n$, then BAB' is nonnegative definite.

If A is a symmetric matrix and $\lambda_{\max}$ is the maximum eigenvalue of A, then $\lambda_{\max}I - A$ is nonnegative definite.

20. Norms of vectors and matrices. The norm of a vector x, written $\|x\|$, is a measure of the size or length of x. There is no unique definition, but the following postulates must be satisfied.

1. $\|x\| \geq 0$ for all x with equality if and only if $x = 0$.
2. $\|ax\| = |a|\,\|x\|$ for any scalar and for all x.
3. $\|x + y\| \leq \|x\| + \|y\|$ for all x and y.

If $x = (x_1, x_2, \ldots, x_n)$, three common norms are

$$\|x\| = [\sum_{i=1}^{n} x_i^2]^{1/2}, \quad \|x\| = \max_i |x_i| \quad \text{and} \quad \|x\| = \sum_{i=1}^{n} |x_i|.$$

The Schwartz inequality states that $|x'y| \leq \|x\|\,\|y\|$ for arbitrary x and y, with equality if and only if $x = \lambda y$ for some scalar λ.

The norm of an $m \times n$ matrix A is defined in terms of an associated vector norm by

$$\|A\| = \max_{\|x\|=1} \|Ax\|.$$

The particular vector norm used must be settled to fix the matrix norm.

Corresponding to the three vector norms listed, the matrix norms become, respectively, $[\lambda_{\max}(A'A)]^{1/2}$, $\max_i (\sum_{j=1}^{n} |a_{ij}|)$ and $\max_j (\sum_{i=1}^{n} |a_{ij}|)$. Important properties of matrix norms are

$$\|Ax\| \leq \|A\| \|x\|, \quad \|A + B\| \leq \|A\| + \|B\|,$$

and

$$\|AB\| \leq \|A\| \|B\|.$$

21. Linear matrix equations. If A, B, and C are known matrices, of dimension $n \times n$, $m \times m$, and $n \times m$, respectively, we can form the following equation for an unknown $n \times m$ matrix X:

$$AX + XB + C = 0.$$

This equation is merely a condensed way of writing a set of mn simultaneous equations for the entries of X. It is solvable to yield a unique X if and only if $\lambda_i(A) + \lambda_j(B) \neq 0$ for any i and j—i.e., the sum of any eigenvalue of A and any eigenvalue of B is nonzero.

If C is positive definite and $A = B'$, the *lemma of Lyapunov* states that X is positive definite symmetric if and only if all eigenvalues of B have negative real parts. If $C = DD'$ for some D, and $A = B'$, the lemma of Lyapunov states that X is nonnegative definite symmetric, and nonsingular if and only if $[A, D]$ is completely controllable.

22. Common differential equations involving matrices. The equation

$$\frac{d}{dt}x(t) = A(t)x(t) \quad x(t_0) = x_0$$

commonly occurs in system theory. Here, A is $n \times n$, and x is an n vector. If A is constant, the solution is

$$x(t) = \exp[A(t - t_0)]x_0.$$

If A is not constant, the solution is expressible in terms of the the solution of

$$\frac{dX(t)}{dt} = A(t)X(t) \quad X(t_0) = I,$$

where now X is an $n \times n$ matrix. The solution of this equation cannot normally be computed analytically, but is denoted by the *transition matrix* $\Phi(t, t_0)$, which has the properties

$$\Phi(t_0, t_0) = I \quad \Phi(t_2, t_1)\Phi(t_1, t_0) = \Phi(t_2, t_0)$$

and

$$\Phi(t, t_0)\Phi(t_0, t) = I.$$

The vector differential equation has solution

$$x(t) = \Phi(t, t_0)x_0.$$

The solution of

$$\frac{dx(t)}{dt} = A(t)x(t) + B(t)u(t) \quad x(t_0) = x_0$$

where $u(t)$ is a forcing term is

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, \tau)B(\tau)u(\tau)\, d\tau.$$

The matrix differential equation

$$\frac{dX}{dt} = AX + XB + C(t) \quad X(t_0) = X_0$$

also occurs commonly. With A and B constant, the solution of this equation may be written as

$$X(t) = \exp[A(t - t_0)]X_0 \exp[B(t - t_0)]$$
$$+ \int_{t_0}^{t} \exp[A(t - \tau)]C(\tau) \exp[B(t - \tau)]\, d\tau.$$

A similar result holds when A and B are not constant.

23. Several manipulative devices. Let $f(A)$ be a function of A such that

$$f(A) = \sum_{i=0}^{\infty} a_i A^i \qquad (a_i \text{ is constant}).$$

[In other words, $f(z)$, where z is a scalar, is analytic.] Then,

$$T^{-1}f(A)T = f(T^{-1}AT).$$

This identity suggests one technique for computing $f(A)$, if A is diagonalizable. Choose T so that $T^{-1}AT$ is diagonal. Then $f(T^{-1}AT)$ is readily computed, and $f(A)$ is given by $Tf(T^{-1}AT)T^{-1}$. It also follows from this identity that the eigenvalues of $f(A)$ are $f(\lambda_i)$ where λ_i are eigenvalues of A; the eigenvectors of A and $f(A)$ are the same.

For n vectors x and y, and A any $n \times n$ matrix, the following trivial identity is often useful:

$$x'Ay = y'A'x.$$

If A is $n \times m$, B is $m \times n$, I_m denotes the $m \times m$ unit matrix, and I_n the $n \times n$ unit matrix, then

$$|I_n + AB| = |I_m + BA|.$$

If A is a column vector a and B a row vector b', then this implies

$$|I + ab'| = 1 + b'a.$$

Next, if A is nonsingular and a matrix function of time, then

$$\frac{d}{dt}[A^{-1}(t)] = -A^{-1}\frac{dA}{dt}A^{-1}.$$

(This follows by differentiating $AA^{-1} = I$.)

If F is $n \times n$, G and K are $n \times r$, the following identity holds:

$$[I + K'(sI - F)^{-1}G]^{-1} = I - K'(sI - F + GK')^{-1}G.$$

Finally, if P is an $n \times n$ symmetric matrix, we note the value of grad$(x'Px)$, often written just $(\partial/\partial x)(x'Px)$, where the use of the partial derivative occurs since P may depend on another variable, such as time. As may be easily checked by writing each side in full,

$$\frac{\partial}{\partial x}(x'Px) = 2Px.$$

REFERENCES

[1] Gantmacher, F. R., *The Theory of Matrices*, Vols. 1 and 2, Chelsea Publishing Co., New York, 1959.

[2] Fadeeva, V. N., *Computational Methods in Linear Algebra*, Dover, New York, 1959.

APPENDIX B

BRIEF REVIEW OF SEVERAL MAJOR RESULTS OF LINEAR SYSTEM THEORY

This appendix provides a summary of several facts of linear system theory. A basic familiarity is, however, assumed. Source material may be found in, e.g., [1] through [3], whereas [4] deals with a special topic—the determination of state-space equations, given a transfer function matrix.

1. Passage from state-space equations to transfer function matrix. In system theory, the equations

$$\dot{x} = Fx + Gu$$
$$y = H'x$$

frequently occur. The Laplace transform may be applied in the same manner as to scalar equations to yield

$$sX(s) = FX(s) + X(0) + GU(s)$$
$$Y(s) = H'X(s)$$

whence

$$Y(s) = H'(sI - F)^{-1}GU(s)$$

with $X(0) = 0$. The transfer function matrix relating $U(s)$ to $Y(s)$ is $H'(sI - F)^{-1}G$.

2. Conditions for complete controllability and observability. A pair of constant matrices $[F, G]$ with F $n \times n$ and G $n \times r$ is termed completely controllable if the following equivalent conditions hold:

1. Rank $[G\ FG \cdots F^{n-1}G] = n$.
2. $w'e^{Ft}G = 0$ for all t implies $w = 0$.
3. $\int_0^T e^{Ft}GG'e^{F't}\,dt$ is positive definite for all $T > 0$.
4. There exists an $n \times r$ matrix K such that the eigenvalues of $F + GK'$ can take on arbitrary prescribed values.
5. Given the system $\dot{x} = Fx + Gu$, arbitrary states x_0, x_1, and arbitrary times t_0, t_1, with $t_0 < t_1$, there exists a control taking the system from state x_0 at t_0 to state x_1 at t_1. [In contrast to (1) through (4), this is also valid for time-varying F and G if $t_0 = t_0(t_1, x_0, x_1)$.]

A pair of matrices $[F, H]$ with F $n \times n$ and H $n \times r$ is termed completely observable if $[F', H]$ is completely controllable.

3. Passage from transfer function matrix to state space equations—the Ho algorithm. The determination of state-space equations corresponding to a transfer function, as distinct from transfer function matrix, is straightforward. Given a transfer function

$$W(s) = \frac{b_n s^{n-1} + b_{n-1}s^{n-2} + \cdots + b_1}{s^n + a_n s^{n-1} + \cdots + a_1}$$

state-space equations $\dot{x} = Fx + gu$, $y = h'x$ yield the same transfer function relating $U(s)$ to $Y(s)$ if

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & 1 \\ -a_1 & -a_2 & -a_3 & \cdot & \cdot & -a_n \end{bmatrix} \qquad g = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \qquad h = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$

or

$$F = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & -a_1 \\ 1 & 0 & \cdot & \cdot & \cdot & -a_2 \\ 0 & 1 & \cdot & \cdot & \cdot & -a_3 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & -a_n \end{bmatrix} \qquad g = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix} \qquad h = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

These formulas are valid irrespective of whether the numerator and denominator of $W(s)$ have common factors. The first pair of F and g is com-

pletely controllable, and there exists a coordinate basis transformation taking any other set of F, g and h which are completely controllable to the prescribed form. The second pair of F and h is completely observable, and there exists a coordinate basis transformation taking any other completely observable set to the prescribed form.

If the numerator and denominator of $W(s)$ have no common factor, both sets are simultaneously completely controllable and observable.

When $W(s)$ is a matrix—say, $p \times m$—the algorithm due to Ho [4] provides a convenient route to determining matrices F, G, and H. First, $W(s)$ is assumed to be zero at $s = \infty$. It is then expanded as

$$W(s) = \frac{A_1}{s} + \frac{A_2}{s^2} + \frac{A_3}{s^3} + \cdots$$

where the A_i are termed *Markov matrices.* Then the A_i are arranged to form *Hankel* matrices H_N as follows:

$$H_N = \begin{bmatrix} A_1 & A_2 & \cdot & \cdot & \cdot & A_N \\ A_2 & A_3 & \cdot & \cdot & \cdot & A_{N+1} \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ A_N & A_{N+1} & \cdot & \cdot & \cdot & A_{2N-1} \end{bmatrix}.$$

The next step requires the checking of the ranks of H_N for different N, to determine the first integer r such that rank $H_r =$ rank $H_{r+1} =$ rank $H_{r+2} = \cdots$. If $W(s)$ is rational, there always exists such an r. Then nonsingular matrices P and Q are found so that

$$PH_rQ = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$$

where $n =$ rank H_r. The following matrices "realize" $W(s)$, in the sense that $W(s) = H'(sI - F)^{-1}G$:

$$\begin{aligned} G &= n \times m \text{ top left corner of } PH_r \\ H' &= p \times n \text{ top left corner of } H_rQ \\ F &= n \times n \text{ top left corner of } P(\sigma H_r)Q \end{aligned}$$

where

$$\sigma H_r = \begin{bmatrix} A_2 & A_3 & \cdot & \cdot & \cdot & A_{r+1} \\ A_3 & A_4 & \cdot & \cdot & \cdot & A_{r+2} \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ A_{r+1} & A_{r+2} & \cdot & \cdot & \cdot & A_{2r} \end{bmatrix}$$

Moreover, $[F, G]$ is completely controllable and $[F, H]$ is completely observable.

4. Minimality. If a transfer function matrix $W(s)$ is related to a matrix triple F, G, H by

$$W(s) = H'(sI - F)^{-1}G,$$

then F has minimal dimension if and only if $[F, G]$ is completely controllable and $[F, H]$ is completely observable. The triple F, G, H is termed a minimal realization of $W(s)$. Given two minimal realizations of $W(s)$—call them F_1, G_1, H_1 and F_2, G_2, H_2—there always exists a nonsingular T such that

$$TF_1T^{-1} = F_2 \qquad TG_1 = G_2 \qquad (T^{-1})'H_1 = H_2.$$

REFERENCES

[1] Ogata, K., *State Space Analysis of Control Systems*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1967.

[2] Zadeh, L. A., and C. A. Desoer, *Linear System Theory*, McGraw-Hill Book Company, New York, 1963.

[3] Schwarz, R. J., and B. Friedland, *Linear Systems*, McGraw-Hill Book Company, New York, 1965.

[4] Ho, B-L, and R. E. Kalman, "Effective Construction of Linear State-Variable Models from Input/Output Functions," *Regelungstechnik*, Vol. 14, No. 12, 1966, pp. 545–548.

INDEX